DOCUMENT RESUME

ED 061 084	SE 013 517
TITLE	Manuals on Training of Statistical Personnel at the Primary and Intermediate Levels.
INSTITUTION	United Nations Economic Commission for Asia and the Far East, Bangkok (Thailand).
PUB DATE	63
NOTE	152p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	*Course Descriptions; Data Processing; Guides; Inservice Courses; *Statistical Analysis; *Training; *Vocational Education

ABSTRACT

This manual is designed for organizers and teachers concerned with training personnel working at the primary and intermediate levels in government statistical offices. The major content consists of outlines for nine training courses: (1) Principles and methods of statistics; (2) Designing a statistical investigation; (3) Data collection by field work; (4) Data collection from records; (5) Data processing; (6) Presentation and publication of statistics; (7) Utilization of statistics; (8) Supervision of statistical work; and (9) Statistical organization, co-ordination and standards. A timetable is suggested for each course; then, under each topic (26 in all), an elaboration of the subject matter is followed by discussion of the most suitable teaching methods and materials for that topic. It is hoped that use of such sources will eventually lead to an improvement in the statistics available from developing countries. (MM)

ED 061084

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

MANUAL
on
TRAINING OF STATISTICAL PERSONNEL
at the
PRIMARY AND INTERMEDIATE LEVELS

UNITED NATIONS

SE 013 517

ECONOMIC COMMISSION
FOR ASIA AND THE FAR EAST
Bangkok, Thailand

MANUAL
on
TRAINING OF STATISTICAL PERSONNEL
at the
PRIMARY AND INTERMEDIATE LEVELS

UNITED NATIONS
New York, 1963

E/CN.11/631

UNITED NATIONS PUBLICATION

Sales No.: 63.II.F.8

Price: US$1.25 or equivalent in other currencies

CONTENTS

Page

INTRODUCTION

1. Background 1
2. Scope of the Manual 2
3. Scheme of the Manual 3
4. Choice of courses and topics 3
5. Method of training 5
6. Programming the training 10
7. Organization of training 11

COURSE 1.

Principles and Methods of Statistics 13
- 1.1 Scope of statistics 14
- 1.2 Statistics and their uses 16
- 1.3 Basic concepts—1 20
- 1.4 Basic concepts—2 25
- 1.5 Statistical operations 33
- 1.6 Elementary statistical methods—1 37
- 1.7 Elementary statistical methods—2 42
- 1.8 Elementary statistical methods—3 48
- 1.9 Qualities of good statistics 53

COURSE 2.

Designing a Statistical Investigation 57
- 2.1 Principles of designing 58
- 2.2 Details of designing 63

COURSE 3.

Data Collection by Field Work 68
- 3.1 Field workers' requirements 69
- 3.2 Field procedures 72

COURSE 4.

Data Collection from Records 78
- 4.1 Material for compilation 79
- 4.2 Compilation schemes 82

Page

COURSE 5.

Data Processing 86
5.1 Processing operations and principles 87
5.2 Computation techniques 92

COURSE 6.

Presentation and Publication of Statistics 99
6.1 Tabular presentation 100
6.2 Graphic presentation 104
6.3 Textual presentation and publication 108

COURSE 7.

Utilization of Statistics 113
7.1 Some guides to utilization 114
7.2 Some methods in utilization 117

COURSE 8.

Supervision of Statistical Work 121
8.1 Supervision in general 122
8.2 Supervision of operations 124

COURSE 9.

Statistical Organization, Co-ordination and Standards 129
9.1 Statistical systems 130
9.2 Statistical co-ordination and standards 132

Appendix 1. Project Work 136

Appendix 2. Equipment and Training Material 139

Appendix 3. Publications for Reference 142

FOREWORD

The training of statistical personnel has been a priority problem engaging the attention of ECAFE for the last few years. Shortage of trained staff continues to be a serious obstacle to the development of statistics in many countries of Asia and the Far East, which are making a determined effort towards planned economic and social development and need more and better statistics than they obtain at present for framing their policies and for formulating, implementing and evaluating their plans. There has been general agreement that, to collect and obtain the necessary statistics, an all-round expansion of statistical training programmes of the countries of the region is urgently needed. However, keeping in view the various difficulties involved in overcoming the shortage of trained statistical personnel in a short time, it has been generally recognized that a beginning should be made by strengthening the existing training facilities and improving the quality of existing statistical services, especially at the primary and intermediate levels.

Statistical activity requires trained personnel of various types and at various levels such as administrative statisticians, mathematical statisticians, investigators, analysts, computers, and machine operators. At the higher professional level, trained personnel will be required in relatively small numbers. Training at that level is mostly undertaken by universities and statistical institutes. In some of the countries, e.g., India, Indonesia, Japan, Pakistan and the Philippines, efforts are being made to improve or develop statistical education in universities and other institutes. Arrangements also exist under international programmes of technical assistance for having statistical personnel trained in the more advanced countries. International statistical centres operating in the region, some of which are permanent and others are specialized and of short duration have also been of much assistance in the training of professional statisticians. In the present situation, however, the training of the large number of statistical personnel at the primary and intermediate levels seems to be by far the most important and pressing problem in the development of statistics. The only way to train personnel in these levels is to organize within each country national training centres capable of carrying out suitable in-service and other training programmes. Their establishment would involve the training of those responsible for organizing and maintaining them and may require external assistance in the form of experts, equipment and so on. In this respect there is considerable scope for co-operation among the countries of the region.

ECAFE had been considering ways and means of meeting the inadequacy of statistical personnel. The United Nations Statistical Commission, at its eleventh session in May 1960, had stressed the need for expanded statistical training programmes in the region. The Conference of Asian Statisticians (COAS) has been considering, since its inception in 1957, measures for improving the training facilities. A Working Group on Training of Statisticians formed in 1960 under the auspices of COAS, made a detailed study of the training needs and problems and gave its recommendations. The Working Group suggested, among other things, preparation of a training manual, organization of a regional centre for training organizers of national training courses and technical assistance to countries in establishing training centres. The Commission has endorsed these recommendations and the present Manual on Training of Statistical Personnel at the Primary and Intermediate Levels has been prepared in pursuance of these decisions. This Manual, designed for use by organizers and teachers, makes a departure from the usual textbooks, and aims at imparting a considerable amount of technical skill to those engaged in day to day statistical operations, without losing sight of the principles involved in the techniques. It will provide the users with some amount of flexibility as it allows for the selection and adaptation of topics to suit different types of personnel. The use of a common manual will also achieve a certain degree of uniformity in training in different countries. It is hoped that it will serve as a useful guide for the organizers and teachers at the training centres. Arrangements are being made to supplement the Manual with necessary training notes. A Regional Statistical Adviser has been appointed at ECAFE, and his services can be drawn upon by countries, in matters concerning statistical training.

U Nyun
Executive Secretary

Bangkok
20 August 1963

87

INTRODUCTION

1. Background

1. The survey of statistical manpower and their training has been one of the important projects of ECAFE and has been a standing item in the agenda of the Conference of Asian Statisticians (COAS) established in 1957. COAS has been considering at its different sessions, ways and means of improving and strengthening the training of statisticians in the countries of the region. At its third session held in April 1960, COAS noted that there was an urgent need to train, within each country, the large body of statistical personnel at intermediate and lower levels. The issue was subsequently followed up by a Working Group on Training of Statisticians. In their report (E/CN.11/545 of 28 December 1960), the Working Group has made a review of the statistical needs and training facilities in the countries of the region and has stressed the need for organizing training centres as a means to provide facilities for improving the quality of statistical manpower and of statistics. The report gives priority to the training of personnel at the primary and intermediate levels, and considers in detail the substantive and organizational aspects of the training of such personnel. The report also lays down the requirements of the training and provides an annotated syllabus.

2. The keynote of the recommendations in the report is that the training should not aim only at imparting theoretical knowledge, but should have a definite practical orientation with a view to developing in the personnel at the primary and intermediate levels the necessary skills in various types of statistical work. The recommendations further recognize that even personnel at the primary level should have a broad idea of the purposes for which statistics are produced and the uses to which they are put.

3. Primary personnel would ordinarily include those concerned with items of work such as collection of data, scrutiny and computation. Training intended for such personnel may also be adapted to suit village accountants or other personnel engaged in maintaining basic records. Intermediate personnel would include those engaged in supervising primary staff or assisting professional statisticians and administrative officers. Some of those in the intermediate level may be graduates who would later on become statisticians.

4. It has further been observed in the report that, even for such personnel, training courses of a wide variety may be found necessary, as for instance, a general training on important aspects of statistical work, a special training for a particular census or sample survey, a special training in statistics in a subject field, or a special training in a particular phase of statistical work

such as computation or field work. But taking into account the facts that, in many countries, the same personnel are often required to do different types of statistical work and that full understanding of a particular phase is hardly obtained without sound knowledge of related phases, it has been recommended that a general training which would necessarily contain the basic elements for any special purpose training as well, and which would suit in-service training programmes, be given first consideration.

5. The Working Group has also recommended in the report, the preparation of a training manual on the basis of the syllabus, which would be useful in conducting general statistical courses, and could be adapted to statistical training in particular subject fields. This Manual has been prepared accordingly.

2. Scope of the Manual

6. The Manual is designed for organizers and teachers concerned with training statistical personnel working at the primary and intermediate levels in government offices. Importance is, therefore, attached to governmental and other large-scale statistical activities and not to those relating to specialized scientific work. One of the main objectives of the training envisaged being the development and improvement of the quality of national statistics, the training should have a close bearing on national statistical activities and national statistical series.

7. While the Manual provides outlines of the subject matter for training and gives indications of the method of training, classified under suitable topics, the task of adapting it in order to serve the special needs of each country and the special requirements of trainees attending particular sessions has, inevitably, to be taken up by the respective organizers and teachers. This task can be considerably simplified if training courses are organized within statistical offices or other institutions affording trainees direct contact with statistical work in progress, and if training material for illustration and for exercises are drawn from the national statistical projects which are closest to the trainees' fields of activity.

8. The Manual provides for comprehensive training, within certain time limits, covering a succession of topics. Although for full training at the level visualized all the topics will have to be gone through, it will be possible to make a selection of or to condense or expand the topics, or to make other modifications that may be found necessary to suit the personnel to be trained.

9. The type of training covered by this Manual will also be found useful for training statistical personnel engaged in institutions other than government offices or for training prospective employees for such institutions and offices.

3. Scheme of the Manual

10. The Manual devotes a section to each of nine courses and contains three appendices. Each course is further subdivided into two, three or more topics. At the beginning of each course, a course summary is given, mentioning the topics forming part of the course and the time proposed for instruction by way of lectures and discussions, demonstrations, exercises and visits for each topic. There are twenty-six topics in all, under the nine courses. Under each topic, first an elaboration of its subject matter is given. This would determine the coverage of the training and provide the details around which training can be given. Within each topic there is a second section indicating the mode of training on the topic. This section would give the main subjects for lectures and discussions, the kind of demonstrations that may be held, the type of exercises that may be set, the visits that may be arranged and the training material required, under each topic.

11. A training summary showing the different courses, the topics under them, and the time that may be devoted to the different elements of training is given in table 1.

4. Choice of courses and topics

12. Each of the nine courses could be treated as an independent course. Course 1 deals with general principles and methods, whereas the other courses are concerned with details of statistical operations. Accordingly, Course 1 may be found necessary before any one of the other courses is taken up. For a general training covering all important aspects of statistical work, all the Courses 1 to 9 should be completed by intermediate personnel and all the Courses 1 to 7 by the primary personnel, preferably in the same sequence in which they are listed in table 1.

13. *Intermediate personnel.* Both the primary and intermediate personnel would undergo the same training, as far as a major portion of the syllabus is concerned. Courses 8 and 9 will concern only personnel at the intermediate level. The topics as well as demonstrations, exercises, discussion and visits within a topic which need concern the intermediate personnel only and not the primary staff are indicated separately, at appropriate places. It is important that the intermediate personnel also undergo that part of the training given to the primary staff, unless they already have previous training or possess considerable experience, because such training is necessary to enable them to appreciate the details of work that they may have to supervise, or sometimes do by themselves, later on.

14. *Personnel seeking specialization.* Some of the courses might be intensified if need be, to meet the special needs of personnel who seek only specialized training. Thus, for personnel mainly engaged in the data processing phase of statistical work, Course 1 and Course 5 are of special

Table 1.

SUMMARY OF TRAINING

Course (1)	Topics under the course (2)	Number of hours proposed for instruction					Stages of project work connected with course (8)
		Lecture/ discussion (3)	Demonstration (4)	Exercise (5)	Visit (6)	Total (7)	
1. Principles and methods of statistics	1.1 to 1.9	31 (3)	10 (4)	18 (4)	—	59 (11)	1 to 2
2. Designing a statistical investigation	2.1 & 2.2	14	6	6	—	26	3 to 5
3. Data collection by field work	3.1 & 3.2	16	8	12	4	40	6 to 9
4. Data collection from records	4.1 & 4.2	12	4	6	—	22	10 to 11
5. Data processing	5.1 & 5.2	20 (6)	6	30 (14)	4	60 (20)	12 to 13
6. Presentation and publication of statistics	6.1 to 6.3	18	4	16	—	38	14 to 17
7. Utilization of statistics	7.1 & 7.2	10 (6)	—	10 (8)	—	20 (14)	18 to 20
8. Supervision of statistical work (for intermediate level only)	8.1 & 8.2	(8)	—	(6)	(6)	(20)	
9. Statistical organization, co-ordination and standards (for intermediate level only)	9.1 & 9.2	(12)	—	(4)	(4)	(20)	
	TOTAL:	141 (35)	38 (4)	108 (36)	18 (10)	305 (85)	Project hours: 40

(The figures in brackets show hours to be spent by intermediate level personnel only, and already form part of the figures, if any, to which they are suffixed.)

interest; for primary personnel engaged in compilation of routine statistics, Courses 1, 4, 5 and 6; for primary personnel concerned with field work, Courses 1, 2 and 3; for primary personnel engaged in analysis of data, Courses 1 and 5; for intermediate personnel engaged in analysis of data, Courses 1, 5 and 7; for supervisors (intermediate) who already have some previous training, Course 8 and/or 9, and so on.

15. However, these courses will have to be suitably expanded or supplemented when the training requirement is in statistics relating to a subject

field such as trade, transport or industry or when the training needed is in connection with a particular investigation such as an agricultural census or sample survey. But, for training on topics such as sample survey theory, or econometric methods, it is desirable that they are designed independently, although training on some of the topics in the Manual may usefully be taken up first.

5. Method of training

16. *In view of the emphasis on practical character of the training, instruction is intended to be mainly through demonstrations, assignments and discussions, lectures being reduced to a minimum.*

17. Each of the topics into which a course has been divided will form a convenient unit of instruction. Some of the topics concern principles, while the rest deal with operations and techniques. The time devoted to a topic relating to operations will be more than on a topic dealing with principles. Each topic will be introduced briefly through one lecture. A *lecture note* on the topic, prepared in advance, elaborating the subject matter of the lecture and providing illustrations will then be distributed to the trainees. The subject matter given under each topic in this Manual, will serve as a basis for the preparation of the lecture note. The matter may be expanded by further explanation, by furnishing more examples and by attaching specimens for illustration. The content of each such lecture note may thus be much more than what will be conveyed through one lecture. The lecture note should constitute the main reading and study material for the trainees. After sufficient time has been allowed for its study, one or more lectures followed by discussions on the topic may be held. During such lectures and subsequent discussion, a number of questions for clarification will be raised and answered. The results of each discussion may be summarized and distributed as a supplementary *discussion note* for study.

18. Associated with a topic there may be one or more selected demonstrations, which may be held in the classroom or field as necessary. So also under most of the topics there are selected practical exercises which will be worked out by the trainees in the classroom or in the field. Sufficient explanation will be provided during demonstrations and practical exercises, but these may be followed by discussions as well. Discussions following lectures, demonstrations and exercises will serve as an effective means for co-ordinating and strengthening instruction. During lectures, demonstrations and particularly discussions, copies of specimen sheets and other training material should be made available to each trainee, wherever possible, for simultaneous reference.

19. Besides, a few planned visits to statistical offices or other places of statistical activity are associated with some of the topics. For the benefit of those trainees who will be undergoing all or most of the courses, a

small scale statistical project has been proposed, which the trainees will carry out by themselves through its different stages (appendix 1). Although this project may function independently of training on different topics, the different stages of project work could proceed as a parallel activity, as shown in column 8 of table 1.

20. *Lectures.* The duration of each lecture is taken to be one hour. It will be seen from table 1 that the time to be spent on lectures and discussions together, is less than half the total training time. Since a large portion of the time will have to be devoted to discussions, the number of lectures should be kept small. This will be in keeping with the principle that lectures cannot be the main form of instruction in a training intended to create practical knowledge and skill. In the first place, the lecture method establishes a one way communication only between the instructor and the trainee and would call for a great degree of comprehension and imagination on the part of the trainee to translate the contents of the lecture into working methods and actions. Lectures, moreover, cannot take the trainee a long way in the matter of gaining practical knowledge as perhaps demonstrations and discussions would do. Further, development of skill will be possible only by the trainee doing pieces of work himself and, in the limited time available for the training, he should be helped to gather as much working experience as possible by observation and by doing exercises and participating in project work. An excess of lectures would therefore remove the emphasis on imparting practical knowledge and skill. Lectures are however necessary for presenting subject matter systematically and in a logical and correct setting.

21. To limit the number of lectures to a few per topic will require careful planning. The fact that there will be only a few lectures per topic should not stand in the way of the lecture note prepared on the topic being sufficiently elaborate. The lecture note which is intended for subsequent study by all trainees would incidentally take special care of those trainees who are accustomed to learn more effectively by reading than by listening and observation.

22. *Demonstrations.* Apart from demonstrations on various aspects of statistical work facilitated by the training being so organized as to provide contact with actual work, and in addition to any demonstrations that may be made during lectures and discussions, a few selected demonstrations are proposed. Such demonstrations will have the advantage of presenting several aspects together, in a convenient way.

23. A demonstration is intended to be of two hours duration. There are 19 demonstrations falling under the first six courses. These demonstrations are described separately under each topic in this Manual. Each demonstration will have to be planned in advance so that it can be completed within the two hour period, permitting some time for clarification of doubts. The planning of each demonstration will include the collection

and organization of the necessary material. For some of the demonstrations, sufficient numbers of each item of demonstration material will have to be secured so that each trainee will have with him one set during the progress of the demonstration. Such material will include lists, maps, schedules, questionnaires, survey calendars, field programmes, punched cards, table layouts, tabulation programmes, graphic presentations, organization charts, photographs, and similar material. Other demonstrations will require recorded tapes, slides, wall charts, calculating machines, mathematical tables and other computational aids and accessories or some specially prepared apparatus. Most of the demonstrations will be held in the classroom, but some will have to be held outside.

24. *Exercises.* A few special exercises are proposed under almost all the topics, to be worked out in addition to any project work or other assignments outside the class hours. There are 54 such exercises in all under the nine courses. The coverage of the exercises is described separately under each topic. Each exercise is to be worked out in a session of two hours. An exercise may contain sub-sections. Most of the exercises will be worked out in the laboratory or classroom; but some, particularly those relating to data collection, will be worked in the field. The laboratory or classroom will be equipped with calculating machines, mathematical and statistical tables and other computational aids and with drawing tables and instruments.

25. Exercise sheets corresponding to the different exercises will have to be prepared in advance, and distributed separately for each exercise, together with the necessary blank sheets (square ruled paper) and graph paper for working it out. Maps, schedules, questionnaires or other material that are to be used in working out the exercises should also be provided. Where necessary, hints for working out exercises should be added after the exercise. The exercise completed should be systematically filed by each trainee to gradually build up a practical work book.

26. At the beginning of each practical session, a briefing of about 15 minutes should be given to the trainees by the instructor, who should mention the principles involved in the exercise and indicate the general plan and important steps for working out the exercise. Thereafter the trainees should be encouraged to think and work out the exercise on their own with the minimum of help from the instructor. The practical work books should be periodically examined by the instructors and comments made.

27. Exercises should be so framed that each can be completed within the two hour period. If an exercise is not completed during the appropriate session, for any reason, by a trainee, he should complete it outside class hours and as early as possible.

28. *Discussions.* As much time as possible out of the time allotted for lectures and discussions on each topic should be devoted for discussions because of their importance in the present type of training. Although

discussions need not be held in sessions of equal duration, those which follow lectures can be arranged in two hour sessions. A number of questions arising within the subject matter contained in each topic should be raised and suitable answers or clarifications provided. The questions should naturally be geared to impart practical knowledge, and should be mostly concerned with bringing out problems, points and other details that will be met in practice and ways of tackling them. Some of the questions should however involve clarification of principles as well.

29. Discussions may have to be stimulated. If it is left to the trainees always to take the initiative in asking questions, it is likely that very few questions will be asked and that only a small section of the trainees will come forward with questions. On the contrary, if all the questions are posed and answered by the instructors, the discussion will take the shape of a lecture. It is therefore very necessary to have the discussion preplanned and conducted in an organized fashion. A series of questions and answers within each topic should be prepared by the instructors in advance, but not distributed before or during the session. These questions and other questions as well should be posed to the whole class and trainees asked individually to provide answers. These answers should be reviewed briefly by discussion and the correct answer formulated by the instructors. Discussion will not only promote understanding by the trainees of principles and techniques, but also enable the instructor to judge the grasp on the subject that each trainee has acquired.

30. The questions and answers prepared in advance will enable the instructors to conduct the discussion systematically and with proper utilization of time. They will also form a substantial part of the notes prepared and distributed after the discussion (paragraph 17 above). The notes should include also some questions to be tackled by the trainees themselves. Those relating to each discussion should be distributed as soon as possible afterwards. Questions and answers and questions for answering prepared in advance under each topic, may be utilized for initiating discussions and for the preparation of discussion notes. One or more discussion notes will have to be issued corresponding to each topic. The discussion notes together with the lecture notes on each topic will form the main study material for the trainees.

31. It is likely that some of the trainees will be found to be unaccustomed to the discussion method. One way of stimulating discussion will be to ask each trainee to prepare and present for answering during discussion a set of questions based on his study of the lecture notes on each topic. Special attention may be paid to them during discussion and they should be encouraged to ask questions and give answers. The discussion habit should be developed in the trainees as early in the course as possible and care should be taken to ensure that all trainees participate in the discussion.

32. Facilities should be provided for the trainees to consult instructors individually. For this, consultation or tutorial hours may be set apart. The arrangement made in respect of the number of consultants, instructors or tutors and the time allowed should be adequate to ensure that all the trainees avail themselves of this facility.

33. *Visits.* If the training is organized to take place within a statistical office which carries out work on a large scale, the trainees will have the opportunity of observing the systematic flow of operations of various kinds and details of actual work. Whether this facility exists or not, a few special visits to statistical offices or other places where field work or other activities are carried out have been proposed. The nine visits proposed are connected with the courses on data collection, data processing, supervision and statistical organization. These visits will have to be made by previous arrangement and it should be possible to demonstrate to the trainees different aspects of work during the visit. The duration of each visit, excluding time for travel, has been taken to be two hours.

34. *Project work.* While the lectures demonstrations, exercises, discussions and visits would cover the entire subject matter of the syllabus, in order to illustrate the interrelations of the different stages of statistical work and to provide trainees with additional working experience, a self-contained, small statistical project should be carried out by the trainees. The emphasis in this project work should be not on the volume of work but on the representative nature of the project and the quality of work. Although the project will be a small one, since the different stages will have to be carried out in sequence, it will be spread over a few weeks. Accordingly, project work will become part of the training only for those who undergo relatively long term training, such as those taking Courses 1 to 7 or Courses 1 to 9. Carrying out the poject work after completion of all the courses, although desirable from some points of view, would mean a considerable extension in the total training time. The project should therefore be carried out, side by side with the training on the different topics. Each trainee should participate in all stages of work in the project as an assignment, either individually or in a team as the stage of work demands. Data collection in the project will have to be done outside the class hours. But other stages of work can be done mostly in the classroom. For this purpose the project work has been divided into 20 stages, each stage involving two hours of classroom or laboratory work. The specification of the type of project to be selected and the 20 stages of work are given in appendix 1.

35. *Publications for reference.* A list of selected publications appears in appendix 3. Most of these are international publications, of the manual, handbook or studies-in-method type. In view of the fact that few of the available text-books provide full treatment of practical details of statistical work, it may not be possible to make wide use of text books for the training. The trainees, while they should have easy access to reference

material and other publications in the library, should give priority to the lecture notes and discussion notes and other training literature distributed during training. Accordingly, assisted reading will not form an important part of the training.

36. *Tests.* In addition to following the progress made by each trainee and judging the performance of each during discussions and in working out assignments, periodic and final examinations may be held to assess the trainees. Conducting examinations will not only provide a basis for the award of training certificates, but will increase the seriousness and alertness of the trainees. The examination procedure may be considerably simplified by setting objective questions, whether the test is written, oral or a field exercise.

37. If the number of candidates seeking admission to the training courses is more than the in-take capacity of the training centre, screening tests for admission will have to be held. One of the main considerations in selecting a candidate for training will be his usefulness after training. The objective type of tests will be found useful for the purpose of selection tests. Tests may be conducted in respect of knowledge in subject fields and aptitude in numerical work and/or field work. The knowledge of mathematics and economics, required for the training will be high school algebra and geometry and familiarity with some of the elementary terms in economics.

38. *Orientation lectures.* For the benefit of trainees who join the training centre for the first time, one or two orientation lectures may be arranged, to enable them to adjust themselves quickly to the environment of the training centre.

6. Programming the training

39. It will be seen from table 1 that the total time for instruction, for all the 9 courses, is 305 hours, out of which 220 hours will be spent on primary personnel and an extra 85 on personnel belonging to the intermediate level. This does not include the 40 hours for project work in the laboratory or class room or the time for data collection outside class hours for project work, or time that will be set apart for consultation by trainees or for tutorial work. If training is on all the 9 courses, and is imparted on a full time basis at the rate of about 25 hours per week, it can be completed in about 3 months. If the training is given on a part time basis, it will have to be spread over a longer period. But it is desirable not to extend even part-time training beyond a period of 6 months.

40. In the preparation of the day to day programme of training classes, the training summary in table 1 and the course summaries given at the beginning of the sections on the different courses will be of help. In taking up courses, topics or items for training, the sequence followed in the Manual may be followed. But from the point of view of uniform

utilization of training equipment and for preparing workable programmes of classes, it may be necessary to take some of the items within a topic out of the sequence indicated. This may not always be possible without seriously affecting the unity of development of the training. Topic 5.2, for instance, is one which may be taken out of the sequence and commenced even from almost the beginning of the training. In order to cover the various items of training within the prescribed time, careful planning of the programme of classes and keeping a watch on the progress of training will be necessary.

7. Organization of training

41. The organizational aspects of training, in terms of teaching staff, location, equipment, materials, and administration, will require special attention. They will be largely determined by the specific purpose of the training, the number of trainees per session and the resources available.

42. As the training is essentially practical in character, involving discussions, demonstrations and exercises, individual attention to trainees will be required. Mainly for this reason, it is recommended that the maximum number of trainees per session should be 30.

43. A list of equipment and material required for training a batch of thirty, is given in appendix 2.

44. *Training staff*. The basic requirements of the staff chosen to impart training should be mastery of subject, skill in work, familiarity with problems to be encountered in statistical operations and ability to suggest measures to tackle them. Some of the training staff may be full time teachers, others part time. There should be one full time staff member in charge of organization, co-ordination and direction, and perhaps two other full time teachers, one each with experience in large scale field work and in data processing respectively. A number of part time teachers with competence in statistical operations in various subject fields will also be needed. They may be drawn from the statistical offices, training institutes and, if necessary, from universities.

45. The training should be located as far as possible in operating statistical offices and organized with close collaboration between as many statistical offices and institutions as possible. It will be possible to utilize a wide range of facilities offered by the various collaborating organizations. Training on different topics may be held at different offices.

46. There may be a number of administrative and instructional problems to which attention has to be paid both in establishing a new training centre and in running it. They are as follows:

(1) *Organization of training centre*

1. Preparation of scheme and budget for centre.
2. Securing funds, accommodation and equipment.
3. Preparation of prospectus, form of application, syllabus and other promotional literature.
4. Enlisting co-operation from other organizations from within and outside the country.
5. Enlisting and training if necessary, teachers and other staff.
6. Preparation of training material.

(2) *Maintenance of centre*

1. Setting up a centre office.
2. Providing library facilities.
3. Transport arrangements (regular, field work).
4. Maintenance of calculating machines and other equipment.
5. Arrangements for issue of stock including laboratory equipment and stationery and for distribution of training material.
6. Arrangements for editing and printing instruction sheets, training notes, test questions.
7. Hostel or other accommodation for trainees and visiting teachers.
8. Circulation of prospectus and other literature, and receiving applications from candidates.
9. Screening of applications (admission tests).
10. Registration of trainees, maintenance of attendance records, progress reports, and examination results and periodic review of training.
11. Preparation of time schedules and training programmes for day to day training.
12. Co-ordinating and supervising instruction.
13. Conducting examinations.
14. Awarding certificates at the end of training (including printing certificates).
15. Preparation of training reports.

Course 1

PRINCIPLES AND METHODS OF STATISTICS

This course is intended to be gone through by trainees both at the primary and intermediate levels, with the exception of one topic meant for those at the latter level only.

The course will provide the necessary background for the eight courses which follow. It could also serve as an independent course on general principles and methods in statistics.

Training will be given through lectures, discussions, demonstrations and exercises. Since the course is designed to provide knowledge of statistical concepts, terminology and principles and to give an introduction to elementary statistical methods, the emphasis in training under this course will not be so much on operational skill as it will be in the subsequent courses. Exercises under this course should be confined to giving the trainee a clear understanding of the principles and techniques through as many brief but representative examples as possible, but not those which involve details of procedure or voluminous material. Numerical work given in this course, unlike that in Course 5, may be simple enough not to require mathematical formulation or the use of calculating machines.

Course summary:

Topic	*Number of hours proposed for instruction*			
	Lecture/ Discussion	*Demonstration*	*Exercise*	*Total*
1.1 Scope of statistics	3	–	–	3
1.2 Statistics and their uses	3	–	2	5
1.3 Basic concepts – 1	3	–	2	5
1.4 Basic concepts – 2	3	2	2	7
1.5 Statistical operations	3	2	–	5
1.6 Elementary statistical methods – 1	5	–	4	9
1.7 Elementary statistical methods – 2	5	–	4	9
1.8 Elementary statistical methods – 3 (for intermediate level only)	(3)	(4)	(4)	(11)
1.9 Qualities of good statistics	3	2	–	5
TOTAL:	31 (3)	10 (4)	18 (4)	59 (11)

(The figures in brackets show hours to be spent by intermediate personnel only and form part of the totals below which they appear.)

TOPIC 1.1

Scope of statistics

1. The statistical approach is part of our everyday thinking. It is this approach by which we gain experience and obtain guidance in many matters of life. In personal matters, we may not formulate our problems in a very scientific manner, we may not make a systematic record of the facts we observe, we may make our computations only mentally, and we may draw conclusions that are clouded by general impressions. Nevertheless, it is the statistical way of thinking that enables us to describe or characterize persons, objects, situations and phenomena with some reliability; to make assessments and comparisons in an objective manner; to make wiser decisions by sifting the true from the untrue, the probable from the improbable, the relevant from the irrelevant; and, by adding facts and findings to our experience, to make decisions and formulate courses of action in such a way that the risks or losses involved are kept at a possible minimum.

In the wider and more important spheres of government and business, in public life and scientific research, statistical techniques and statistical information not only play a magnified role but constitute an organized and systematic methodology, which calls for a good deal of skill and craftsmanship for its successful application.

2. *Role of statistics*

Statistics, as a subject for study and practice, is based on the fact that information is provided in its most efficient form when it is expressed numerically. The importance of the subject lies in the fact that many things and events around us are capable of being described and studied quantitatively. Even many descriptions which are at first qualitative in form can be suitably expressed in numbers. Numerical information can arise in almost all fields of human activity and therefore statistics has a place in every field. For this reason, the subject of statistics remains a body of scientific principles and methods concerned with numerical information, and is not a separate science like economics, biology, physics or chemistry, though statistics has its place in all of them.

Briefly, the role of statistics, in any field of application, is as an aid to solving problems, or to be more specific, to answering a variety of questions raised in solving problems. This function it performs through certain stages. The problem is first suitably formulated and resolved into a number of unambiguous questions; information relevant to each question is collected as accurately as possible; these items of information, otherwise called statistical data, are carefully studied and the results of the study are then utilized to frame an answer to each question.

For example, suppose a country is facing the problem of tackling a famine which has just broken out. Some of the questions that are immediately raised are: What is the number of people for whom food has to be provided? How many are likely to be displaced from their homes? What are the causes which led to the famine? Can it be that hoarding is the main cause? How much time will be taken to set right each of these causes? How much stock of food is available? How much food is to be imported? How much foreign exchange is available for importing food? How much time will it take for imported food to reach the country? What will be the effects of the famine? To what extent are prices likely to rise? Is price control necessary?

The answers to these and similar questions are provided with the help of either statistical information readily available, or information specially collected for each purpose. The type of answer given may differ from question to question. The answers to some of the questions are certain numerical values which can be directly provided. Some of the questions require estimates to be made. Some questions require verification. Others require a study of interrelations of factors before answers can be given. The correctness and reliability of each answer will depend on the amount of information used, its quality, and the efficiency of the methods of using that information to frame the answer.

The questions asked above, are typical of most of the problems which statistics can help to solve. Statistical activities are thus not confined to compilation of numerical facts, but include: formulation of problems, planned collection of information, analysis of the information and utilization of the results of the analysis to meet specific objectives.

3. *Emergence of statistics*

Until about three centuries ago, the subject of statistics comprised only the compilation, for use by the state, of certain basic data such as number of people, number of livestock, or amount of agricultural wealth. The compilation of such information was carried out as an administrative duty. Not much by way of technique, either in the collection of data or in the analysis and interpretation of figures, was known. Statistics had not yet developed as a subject by itself and the term 'statistics' simply meant numerical information. This usage is still in vogue. Such statistics continue, even now, to be compiled on an extensive scale and as an important activity and they form a very wide branch of the subject of statistics.

In the last several decades, many developments have taken place also on the theoretical or analytical aspects of numerical information, particularly in the designing of collection of statistical data, their analysis, and interpretation. These developments have been a result of increased use of statistical information in diverse fields, together with the coming into being of the mathematical theory of probability.

Training on Topic 1.1

4. *Lectures and discussion*

(1) Presentation of contents of sections 1 to 3 above.

(2) Consideration of details of the process of evolution of modern statistics as a scientific methodology and the general role it plays in different fields of application.

5. *Training material*

Lecture notes.

TOPIC 1.2

Statistics and their uses

1. The use of statistics are many and varied. Their use is generally restricted only by the lack of adequate statistical data or due to absence of knowledge as to how to use them correctly and effectively. History shows that, in times of expansion and new activity, in war and emergency, and in reconstruction and planned development, always new types of information are required, and the collection of new statistics, their use and methods of use gain considerable impetus.

Statistics are produced by various agencies at various levels and on various scales. So also there are many types of users of statistics. The producers of statistics create them mostly for their own use; but those statistics are useful for others as well. Accordingly, all users do not produce their own statistics; they may have to depend partly or solely on other sources.

2. *Producers of statistics*

Those who produce statistics may be broadly classified into four categories:

(1) *Individual investigators:* Statistics produced by individual workers are the result of making experiments and other studies, often with a view to discovering scientific facts. Such data arise in the natural and social sciences in fields like genetics, psychology, physics, or in economics in fields like agriculture and industry. These data are of scientific quality and value; they are usually concerned with one or a few specific problems and are produced on a small scale. There are, however, some research organizations which produce statistics in a similar manner, but on a larger scale. These statistics, if published, appear in journals or in reports.

(2) *Individual organizations:* Business organizations like manufacturing firms and retail firms, local authorities like municipal offices, transport organizations like railway or motor transport companies and other such organizations, produce statistics concerning their activities. Such statistics are compiled mostly from records, like payrolls, customer accounts, progress reports, invoices, or supply indents, which accumulate in the course of routine operations. Besides, they may compile statistics based on special inquiries made by them either within the organization or outside. The inquiries may be designed either to answer specific questions from time to time or to provide information once for all.

The statistics produced by these organizations will be available in their statistical documents or in reports. Some of the statistics may be of a confidential nature. Others will be passed on, in self interest or under obligation, to organizations such as newspapers, chambers of commerce, trade associations and the government.

The larger bodies like associations and the government consolidate the statistics received from the different organizations and create a wider set of statistics relating to activities of several similar organizations as a whole. Although, in respect of such statistics, the larger bodies do not produce the data first hand, they are considered as producers of at least secondary statistics, in so far as such statistics are the result of further compilation. Often these secondary statistics are supplemented by statistics based on special inquiries conducted by the larger bodies with a view to filling in gaps in information contained in the original data.

(3) *The government:* By far the greatest producer of statistics in any country is the government. Statistics produced by the government relate to the country's welfare and mostly to the national economy. A considerable part of government statistics are compiled from records which accumulate in day to day administration or from statistical returns received from other sources as provided under law. Other statistics are obtained by conducting special inquiries either periodically or occasionally according to demand for particular types of information. Such special inquiries may be conducted by the government directly or through agencies appointed by the government. Governmental statistics which are not of a confidential nature are published in administrative reports, statistical bulletins, statistical abstracts, statistical year books, census reports, sample survey reports and like publications.

Each country has its own statistical system. The quantity and variety of statistics produced differ from country to country. The major subject fields in which modern governments compile statistics may be classified as follows:—

Population and health
Agriculture and irrigation
Mining, power and industry

Trade, foreign and internal
Transport: railway, road, shipping and civil aviation
Communications
Education and culture
Housing
Labour and employment
Prices
Public finance
Money and banking
Balance of payments
National income, national accounts and financial flows
Planning.

Statistics on subjects like defence, although compiled by governments, are not published or made available for general use.

(4) *International agencies:* The United Nations compiles and publishes statistics in various subject fields, based on data received from individual nations and international organizations. Organizations such as the International Monetary Fund, Food and Agriculture Organization of the United Nations, International Labour Office and the World Health Organization also produce international statistics in their respective fields. These agencies, by developing statistical standards, by contributing to knowledge of statistical methodology and by offering advice, also assist nations to produce statistics that are comparable on an international basis and to develop new statistics where they are lacking.

2. *The government as a user of statistics*

Governments use statistics as an aid in conducting the affairs of the state, or in other words, in administration. Administration in its wider sense involves not only day to day control, but also the making and execution of policies extending over periods of time. Some administrative actions are taken in accordance with established rules, while some are based on decisions dictated by the course of events. Decisions thus have to be made with regard to many individual actions as well as policies of action. It is in relation to decisions behind actions and policies that statistics are most useful. They are also useful in guiding action and evaluating the results of action, thereby allowing changes to be made in decisions, if any. This does not mean that all policies and actions are, or need be, based on statistics; or that, for making a specific decision, all the necessary information will be found in the statistics available. However, often the same set of statistics can be diverted to several uses. The following are some of the government uses of statistics.

(1) *Routine administrative control:* The various types of government statistics describe the nation as a whole or its component sectors, in terms of different characteristics. Statistics describing different entities of a similar type enable comparisons to be made. Thus for instance, comparisons are possible between different geographical areas, different communities or different time periods. Such comparisons provide indicators of differences and indicators of progress. These indicators will show out if something is wrong and serve as signals for action. Examples of taking action on the basis of indicators are:

(*a*) If yield of crop is very low in some areas as compared to others, steps may be taken for providing irrigation, fertilizers or other facilities.

(*b*) If wages fall below subsistence level, enactment for a minimum wage may be introduced.

Statistics as indicators also serve as a check on administrative action. Thus for example, if steel allocated adds up to more than the amount of steel available, evidence is provided that the system of allocation has gone wrong.

(2) *Policy making:* Apart from making decisions and actions on current matters, the government has often to formulate legislative, executive or economic policies, for future action. Examples are: policies of compulsory education and nationalization of industries, budget policy, trade policy, agricultural policy, credit policy, wage policy, taxation policy. The role of statistics in policy making is by summarizing experience of the past and helping to establish results to be expected in future. Firstly, by analyzing statistics in those subject fields closely related to the issue, the policy maker is enabled to select a few alternate policies for consideration. By further analysis of the related statistics he will be able to work out the results that can emanate from the different policies, if adopted. This will help him to choose one reasonable policy out of several under consideration.

Further, when a policy has been in operation for some time, statistics of the results achieved can be used to examine whether the policy has been successful or not. If not, the policy may be revised.

(3) *Planning for economic development:* For accelerating economic growth, many countries adopt planned economic development. In a development plan, certain targets are to be achieved at the end of a prescribed time period. This will involve the framing of a planned policy for achieving each target. Statistics will be useful in framing the plan policies and in determining the magnitudes of the targets corresponding to each policy, if adopted. Thus for planned production of a commodity like steel or textiles, some of the statistics useful for fixing the targets are those relating to production, resources, capacity for expansion and import and production of machinery. For framing the financing policy of the plan, statistics relating to domestic savings and foreign capital will be useful. The financing policy may further involve a taxation policy for mobilizing domestic savings.

4. *The outside user of government statistics*

Outside the government, there are many private users of government statistics. They include business men who require timely information concerning the problems and prospects of their trade. Detailed information will be of interest to them about stocks and prices of raw materials, wage rates, imports and exports, income structure of consumers and similar characteristics related to their trades. There are many other users as well, like economists, statisticians, politicians, historians and journalists.

Training on Topic 1.2

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Illustration of how an organization (a government office, or an industrial establishment) produces its own statistics and makes use of statistics produced by itself and by others.

6. *Exercise*

Example:

(*a*) Matching major national economic characteristics given in one list with minor characteristics in another list.

(*b*) Selecting from specimen documents characteristics which would lead to useful statistics and listing the uses to which the resulting statistics can be put to.

7. *Training material*

Lecture notes. Exercise sheet.

TOPIC 1.3

Basic concepts—1

1. *Statistical data*

A distinction may be made between statistical data and statistics. Statistics are numerical facts derived from statistical data, by processes like rearrangement and condensation. Statistical data can be either qualitative or quantitative, but statistics are always quantitative. (It is possible that statistics derived by one person or for one purpose serve as quantitative statistical data for another.)

Statistical data are the result of observations of characteristics on individuals. The three words "observation", "characteristic" and "individual" are used here with meanings, to some extent, special to statistics. These meanings will be clear from the following examples of making observations leading to statistical data.

1) An agricultural investigator visits farms and by inquiry and using other resources at his disposal, finds out and records for each farm, the area, type of crops, number of livestock and value of implements in use.
2) An inspector in a manufacturing plant takes each bolt produced and, with the help of a micrometer, notes its thickness.
3) An accountant in a firm draws up the balance sheet at the end of a year.
4) An official in a government office compiles the quantity and value of textiles exported by the country in a month from records received from the customs office.

Each of these involves, besides an observation procedure, (*a*) a statistical individual and (*b*) one or more characteristics, which are identified and presented below.

Example	*Statistical individual*	*Characteristic(s)*
1)	Farm	1) Area, 2) Type of crop, 3) Number of livestock and 4) Value of implements
2)	Bolt	Thickness
3)	Firm (assuming that accounts of different firms for the same year are of interest)	1) Receipts, 2) Expenditure and 3) Balance
4)	Month (assuming that exports by the country for different months are of interest)	1) Quantity of textiles exported, 2) Value of textiles exported.

2. *Statistical observation*

1) **The word 'observation' has a wider meaning than usual in statistics.**
2) **Observation is the process by which information from a source is extracted and recorded.**
3) **Observation implies a statistical individual and one or more characteristics of the individual, which are of interest.**

4) The observation method may be different for different characteristics on the same individual.

5) Observation is carried out usually by the human agency with or without the help of instruments.

Among the methods of observation are the following: measurement, counting, personal judgement, taking answers based on interviews, copying from existing records, using self-recording instruments.

6) An observation made by an agency on an individual may be or may not be based on observations made at preceding stages, on the same individual or on component individuals. (Thus, in example 2, the inspector makes a direct observation on the bolt. But, in example 1, the investigator may have to record the area of the farm on the basis of a survey map, or he may have to record the value of the implements on the basis of a price list or accounts shown by the farmer. In example 3, the observation made could be regarded as one made on a compound individual, on the basis of observations made at earlier stages on component individuals, namely, various items of receipts and expenditures of the firm, at different times in the year.)

7) It is important in any problem to decide on what to make the observation, or in other words, what should be the statistical individual.

8) The term 'observation' is sometimes, also used as an abbreviation for 'result of observation (characteristic(s) as observed) on an individual'.

3. *Statistical characteristics*

1) The term 'characteristic' in statistics stands for a feature, property or aspect, in terms of which an entity (whether to be treated as a statistical individual or not) will be described.

2) Several characteristics may be necessary, to describe an entity, for a specified purpose. Example:

Purpose	*Entity*	*Characteristics*
Military recruitment	Male adult person	Height, weight, girth, freedom from diseases, marital status, number of dependents.

3) Characteristics of a broader type may be subdivided into characteristics which can be more easily observed. Example:

Hygienic condition—type of drainage, type of water supply, nearness to factories, accumulation of garbage.

4) A characteristic will vary in its manifestation. In other words, a characteristic can take alternate forms and a particular entity would possess only any one of these alternate forms with respect to each characteristic. Examples:

	Characteristic	*Alternate forms*
a)	Marital status	Single, married, widowed, divorced, separated.
b)	Number of dependents	0, 1, 2, 3, 4, ...

Thus, with respect to the above two characteristics, a person may be described as: married and having 3 dependents,

5) A characteristic of which the alternate forms are numerical values, is called a quantitative characteristic or a *variable.* Examples: income, age, value of grains imported, quantity of electricity produced, household size, number unemployed.

6) A characteristic of which the alternate forms are qualitative categories, is called a qualitative characteristic or an *attribute.* Examples: economic status, sex, educational level, occupation, industry.

7) A variable which can take only certain specific numerical values, is called a discrete variable. Example: a count like number of members in a household or number of motor accidents per day in a town.

8) A variable which is capable of taking *any* numerical value (in some range of values) is called a continuous variable. Examples: age, temperature, value of product.

Note: Even a variable which is inherently continuous, can be observed in practice, only in terms of discrete values, due to limitations in making measurements.

9) The same characteristic may, depending on the context, assume either the form of an attribute or that of a variable. Example: balance of payments—as an attribute would take the alternate forms: favourable, unfavourable; and as a variable would take any positive or negative numerical value.

4. *Statistical individual*

1) The term 'individual' connotes something which will not be further sub-divided in a given context and is the entity on which the observation is made.

2) The individual can be any type of entity, provided it is capable of being identified and observed. Examples: object, group of objects, person, portions of bulk material, geographical area, institution, time period, trade, nation.

3) The individual may be qualified or conditioned by other entities. Examples: farm within a particular geographical area, coal mine relating to a year, worker in a particular industry in a particular locality.

4) The individual may be qualified or conditioned by characteristic manifestations. Examples: male person, two member urban household, hill tract village, re-export commodity, unskilled female labourer.

5) An individual may present one or more characteristics for observation.

6) Observations (of one or more characteristics) on a number of individuals all of the same type, will constitute a separate series of statistical data. Examples: *a*) The export and import figures of a country for different years will form one series of statistical data with the year as the individual, whereas export and import figures for a year, of different countries, will form another series of statistical data with the country as the individual. *b*) Characteristics like family size, monthly expenditure and living room space, observed on the different families in a village, will form one series of statistical data, whereas characteristics like age, sex, educational level and marital status observed on members within the families in the village, will form a separate series of statistical data.

7) The term 'individual' may be used to denote either the general individual class, or a particular member of the class. Thus, when it is stated that the farm is the individual, the class is referred to; but when it is stated that the third individual observed was farm F, a particular farm is referred to.

8) The term 'individual' is also sometimes used as an abbreviation for 'observed characteristic(s) relating to an individual'.

9) It is important to identify, in any statistical data available, the statistical individual concerned, because in the handling of statistical data, information about the individuals such as their type and number will be required.

Training on Topic 1.3

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Explanation of the basic concepts with selected examples.

6. *Exercise*

Example:

Familiarity with observation of characteristics of different types: attribute, count and measurement.

(*a*) Observation of one characteristic of each type (*i*) repeatedly on a single entity, (*ii*) once on each of a number of different entities.

(*b*) Observation of a set of characteristics of different types, all once on each of a number of different entities.

The observations may be made on special objects supplied, or on objects around the classroom. For observations of the measurement type, even lengths of straight lines drawn on paper may be measured.

In this exercise, incidentally, ideas about accuracy attainable with a given measuring instrument and methods of rounding off measurement values, may be brought out.

7. *Training material*

Lecture notes. Exercise sheet.

TOPIC 1.4

Basic concepts—2

1. *Statistical population*

1) The statistical population is a composite entity capable of being defined in terms of its component entities and their characteristics.

2) When the characteristics of an entity can be determined by direct observation on the entity, however large or complex the entity is, the entity is *not* referred to as a population.

Example: If the name or geographic location of a country is of interest, the country is not regarded as a statistical population, rather, in this context it is only an individual.

3) When the characteristics of an entity are such that they can be determined only through an aggregation of the same or other characteristics observed on a number, usually a large number, of statistical individuals, the original entity will be regarded as a statistical population.

Examples: Consider the following characteristics which are to be determined.

(*a*) The cultivated area of a country.

(*b*) The human population (number of inhabitants) of a state.

(*c*) The number of persons dying in a particular region due to cholera, in a year.

Each of these is a population characteristic which can be built up out of observations of the same or other characteristic on all the individuals constituting the population. The observation and aggregation involved in each of the three cases are shown below:

OBSERVATION		*AGGREGATION*	
Individual	*Characteristic*	*Population*	*Characteristic*
a) Plot of land	Area under cultivation	The set of all plots of land in the country	Total area under cultivation in all plots of land in the country
b) Habitation	Number of persons	The set of all habitations in the state	Total number of persons in all the habitations in the state
c) Death (each case of death)	Cause of death (cholera or not)	All death cases in the region, in the year	Number of cholera deaths in the region in the year

Note that in the third example the characteristic of observation is different from the characteristic obtained for the population. In fact, the former is an attribute and the latter a discrete variable.

A statistical population is thus the complete aggregate or totality of statistical individuals in relation to specified characteristics.

4) A population may be divided into sub-populations for organizational or other reasons.

Example: In order to find out the electricity generated during a period of two years, the power generating station may be treated as the individual and the two year period may be sub-divided into months. The set of power generating stations operating in each month will form a sub-population, there being 24 sub-populations in all.

5) The total number of individuals in a population is called the *size* of the population.

6) A population made up of a determinable number of individuals is called a *finite* population. The statistical populations referred to in the foregoing examples, are all *finite* populations. But all populations need not be so.

7) When the characteristic specified is such that observations on an unlimited number of constituent individuals, are admissible, the totality of

such individuals is called an *infinite* population. This totality would remain abstract, but is still defined by the individual and the characteristic(s). Examples: Consider the following questions.

(*a*) By using a certain type of fertilizer how much will be the increase in the yield of rice crops?

(*b*) What proportion of people will die if there is an outbreak of small-pox?

(*c*) What should be the output per week of a particular type of steel plant?

Note that these questions are different from questions like: What was the increase in the yield noted in a particular manurial experiment? or What was the output of the steel plant during the last month? Each of the above questions requires the determination of a characteristic. The determination of the characterisic further requires observation to be made on statistical individuals in each case, as follows:

	Individual	*Characteristic to be observed*
(*a*)	Fertilizer experiment	Yield of rice crop
(*b*)	Small pox outbreak (each occasion)	Number of persons dying
(*c*)	Steel plant	Output per week

Note that the best answer to each of the three questions obviously cannot be given on the basis of observation of the characteristic on one individual, because the characteristic manifestation will differ from individual to individual. Each fertilizer experiment for instance would give a different yield of rice crop.

Consider now how many observations on different individuals, for each characteristic will it be possible to make, in all? This cannot be answered because the number of individuals that can be observed in each case is unlimited. The totality of statistical individuals in each of these situations constitutes an infinite population.

8) The term 'population' is sometimes used also to refer to the 'totality of sets of observed characteristics on individuals'.

2. *Sample*

1) A specified sub-set of statistical individuals of a population is called a sample.

2) A sample will ordinarily consist of more than one individual.

3) The number of statistical individuals in the sample is called the sample *size*.

4) The unit in terms of which a sample is selected is called the *sampling unit*. The sampling unit need not always be the statistical individual.

Example: For the purpose of determining the price of a commodity, prevailing in urban areas of a state, first a sample of the towns may be selected out of all the towns in the state; then from each town selected, a sample of some markets may be chosen and the prices of the commodity in these markets observed. The town and the market are both sampling units. But only the latter is the individual.

5) There are several types of samples, named usually after the manner in which the individuals constituting the sample are selected out of the population. Some types are: equal probability or simple random sample, stratified sample, multi-stage sample, multi-stage stratified sample, cluster sample (Topic 2.1).

6) For the purpose of determining population characteristics, instead of observing the characteristics on all the individuals in the population, the individuals in a sample only may be observed. Then the sample characteristics will be utilized to determine approximately the population characteristics.

7) The error involved in such an approximation is called 'error due to sampling' or *sampling error*. The magnitude of the sampling error will depend mainly on: (*a*) the manner in which the individuals in the sample are selected, (*b*) the size of the sample and (*c*) the extent to which the characteristic manifestations differ from individual to individual in the population.

8) In dealing with only a sample instead of the entire population, there are, however, considerable gains, especially in time and cost, not only in respect of making observations of characteristics but also in the subsequent handling of data.

9) If the population is finite, sampling may or may not be adopted. If the population is very small in size, sampling may not be necessary or desirable.

10) If the population is infinite, only sampling is possible; and sampling error is unavoidable in that case.

11) Generally, the larger the sample size is, the less the sampling error will be, although the time and cost involved will be more. How big the sample from any population should be has, therefore, to be decided in such a way that the harm or damage caused by sampling error plus the time and effort expended in the study of the sample is the minimum possible.

12) It is the practice sometimes to use the term 'sample' also for the sets of observed characteristics on individuals in a sample.

3. *Typical statistical data*

Whatever be the way in which data are recorded at the time of making observation on individuals, it is always possible to draw up a list or series giving the identification of each individual and the characteristics as observed on each of them. (Making such a list is not a necessary step for the handling of data). Given below is a section of such a series relating to some characteristics observed on 1,215 industrial workers. This series represents several aspects of statistical data.

Serial number (1)	*Worker identification number* (2)	*Sex* (3)	*Age (years completed)* (4)	*Marital status* (5)	*Earnings, weekly (dollars)* (6)	*Dependency (number of dependents)* (7)
1	X 0237	male	32	single	67	0
2	X 0313	male	47	married	131	5
3	X 0752	female	39	married	102	1
4	X 1319	male	28	single	89	1
5	X 1615	male	41	widowed	113	3
6	X 1739	female	30	divorced	97	2
.	...	...	..	...	...	.
1214	Z 8321	female	27	single	115	0
1215	Z 9302	male	39	married	119	2

The statistical individual in the data is the industrial worker. Five characteristics have been observed on each worker, namely, sex, age, marital status, earnings and dependency. Each row in the series represents an individual in relation to the five characteristics.

Column (1) shows the serial number of the observation; column (2) the identification of each worker and columns (3) through (7) the five characteristics observed on the individuals.

Of the five characteristics, two, namely, sex and marital status are attributes. Age and earnings are continuous variables and dependency a discrete variable.

Since more than one characteristic are observed, the series constitutes multiple characteristic data.

Some of the population characteristics for the determination of which the data can be used are: (*a*) the number or proportion of male workers, (*b*) the number or proportion of female workers, (*c*) the age group in which the workers fall, (*d*) numbers or proportions of workers in different marital status types, (*e*) total or average earnings of workers, (*f*) total or average number of persons depending on workers.

There are also a number of other population characteristics for which the data can be the basis. For example, (*a*) the age distribution (i.e. the number or proportion of workers, having each of the different ages, or falling within age groups like 20-24, 25-29), (*ii*) the association between sex and earnings, (*iii*) the association between age and earnings, (*iv*) the extent of variation in earnings from worker to worker.

If the 1,215 observations constituting the data are related to all the workers in a factory and if the data were collected for purposes concerning the factory only, the 1,215 individuals form a finite population and the series forms population data.

A subset of rows in the series selected in some specified way would form sample data.

Supposing that, for the purpose of a sociological study, the 1,215 observations were made on workers drawn from various factories belonging to various industries and at different times, to represent the industrial working class in general (including industrial workers of the recent past, of the present and those yet to come in the near future), then the entire data or a part of it may be regarded as forming a sample from an infinite population.

4. *Types of statistical data*

Statistical data can be classified from various points of view and data may be called by different names according to the classification criterion. Thus when data are collected first hand by original observation, as in a special investigation like census or sample survey, they will be called *primary data;* and when data for a purpose are collected by subsequent observation on the results of a previous observation, as in selecting information from published statistics, they will be called *secondary data.*

Some other classifications are listed below:

	Criterion	*Classes*
1)	Coverage with respect to individuals	Population data; Sample data
2)	Type of characteristic	Attributes data; Variables data (Qualitative data; Quantitative data)
3)	Number of characteristics	Single characteristic data; Multiple characteristic data

4) Number of variables	Univariate data; Bivariate data; Multivariate data.
5) Type of individual	Time series (also called chronological or historical data); Geographical data; Other data.

Training on Topic 1.4

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Explanation of the basic concepts with selected examples.

(3) Presentation of examples of primary and secondary data and other types of data.

6. *Demonstration*

Example:

Introduction to characteristics of individual, sample and population. The apparatus required is the following: A closed box containing 300 spherical beads, all of the same size. The beads are so made that (*a*) 60 of them are of one colour (pink) and the other 240 are of another colour (white) (*b*) No black dots appear on 110 of them, and 1, 2, or 3 black dots appear on each of the rest to give the following frequencies:

Number of dots	*Number of beads*
0	110
1	100
2	70
3	20
	300

and (*c*) on each of the beads one of the following values is written: 2.1, 2.2, 2.3, 2.4, 2.5, giving the following frequencies:

Value	*Number of beads*
2.1	10
2.2	40
2.3	200
2.4	40
2.5	10
	300

The 300 beads constitute a population. There are three characteristics specified on the individual bead namely, (1) colour, (2) number of dots and (3) value. Three of the population characteristics, corresponding respectively to these individual characteristics are: (1) proportion of pink beads, (2) average number of dots per bead (or total number of dots), (3) average value per bead (or total of values). (In this case the above population characteristics are already known to be 0.20, 1.0 and 2.3 respectively.)

The bottom of the box is a sliding panel which has 16 depressions on it to receive 16 beads. This panel can be taken out with 16 beads on it, by pushing it out with another similar panel which will replace the former.

The demonstration is started with a brief description of the purpose, but assuming that the population characteristics *are not known.* The beads in the box are mixed well and a sample of 16 beads are taken by sliding the panel. Each of the 16 beads is observed for the three characteristics and the results are systematically recorded for each bead. The sample characteristic is calculated corresponding to each of the three population characteristics. The fact that these are estimates of the population characteristics is pointed out.

Similar sample selection and calculations are repeated 5 times, the sample taken being returned to the box each time and contents mixed well. Mutual differences in the sample estimates are pointed out and are explained in relation to sampling error.

The actual population characteristics, known in this case, are then disclosed and how far the sample characteristics deviate from the actual values is discussed.

In order to show that the sampling error can be less if the sample size is increased, the demonstration may be repeated afresh with two draws of the panel each time to give say 5 samples of size 32. The sample characteristics in these samples would in general be found to deviate from the population characteristics to a lesser extent.

(Note that in this case that sets of 32 beads and not 16 beads are returned to the box each time).

The demonstration will be followed up for the intermediate level personnel at a later stage. (Demonstration 2 under Topic 1.8).

(In the absence of the above apparatus, a tray to contain the beads may be used and the required number of beads selected one by one 'blindly' from different parts of the tray after mixing them well. Also instead of the beads, small cards may be used.)

7. *Exercise*

Example: Acquaintance with statistical characteristics, statistical individuals and statistical data.

(*a*) Indentifying the individuals and characteristics in given statements about observation made and given sets of statistical data.

(*b*) From statistical data consisting of three or four characteristics (to include, attribute, count and measurement) observed on 5 individuals, copying out sample data relating to all possible samples of size 2 in which the same individual does not repeat. Simple calculations of population and sample characteristics from the data.

8. *Training material*

Lecture notes. Exercise sheets. Specimens of different types of data. (section 5 above). Demonstration apparatus (section 6 above).

TOPIC 1.5

Statistical operations

A statistical activity may have several phases involving many operations. Its different phases, whether the activity is a special investigation or a matter of routine compilation, are the following:

(1) Designing, (2) Data collection, (3) Data processing, (4) Presentation of results, and (5) Utilization of results.

The amount of work involved in each of these will depend on various factors such as the number of individuals, the number of characteristics and the extent to which data are subjected to analysis to arrive at population characteristics. The statistical operations involved in these phases are treated in detail in Courses 2 through 7. Only some basic notions are provided below.

2. *Designing*

This phase is concerned with the determination of the nature and size of the activity and with prescribing the conduct of the activity through successives stages and operations. The purpose of designing is to lay down an efficient system of operations which will produce the results aimed at, through the most economic utilization of resources. In designing, therefore, a number of decisions will have to be made in respect of fixing the population and individual characteristics, the individuals and the method of observation, of the number of individuals to be observed, and of the programmes, operational procedures, budgets and administration concerning the different stages.

The various decisions arrived at during designing will be embodied in the instruction manuals concerning the activity. In large scale work particularly, instructions may be prepared for each stage separately.

3. *Data collection*

Broadly there are three methods of data collection, namely:

(1) Field work. Trained field staff visit the spots where the individuals are located, and by adopting suitable methods of observation (Topic 1.3, section 2, (5)) information is collected on specially prepared data sheets.

(2) Receiving statistical returns through transmitting agencies like the post. Prescribed data sheets are completed, by the individuals themselves, or on behalf of the individuals, and received by mail or otherwise.

(3) Selecting information already contained in administrative records, non-statistical returns, or in published or unpublished statistics. The information selected, is copied on suitable data sheets.

In a particular statistical activity any one of the three methods or a combination may be used.

In regard to the data sheet there are two types: (*a*) the schedule or form and (*b*) the questionnaire. These two names are sometimes used indiscriminately to mean data sheet. But a distinction is sometimes maintained between the two. A schedule or form has generally a tabular lay out and information will be recorded on it according to the headings. A questionnaire, applicable when the individual observed or the respondent is a person, contains a series of explicit and direct questions, and information will be entered on it by 'checking' the appropriate answer from among alternatives provided, or in the form of worded answers to the questions. Ordinarily one copy of the data sheet will be used for each individual separately. But in some cases, especially when a schedule or form with a simple lay out is used, information relating to a number of individuals may be accommodated on one copy of the data sheet. A data sheet may run to many pages, depending on the number of characteristics to be observed. In some cases, certain portions of the same data sheet may be of the schedule or form type and other portions of the questionnaire type.

4. *Data processing*

Data processing consists of those operations involved in producing statistics out of statistical data. In many cases, it is concerned with the aggregation of characteristics observed on individuals to obtain sub-population or population characteristics.

The first step in data processing is scrutiny of data sheets with regard to their acceptability and taking action to rectify any defects found. The next step is the arrangement of data. Sorting of the data sheets, in an order different from that in which they are available, will be often necessary. The classification will be according to some of the characteristics or other information recorded on the data sheets. After data sheets have been sorted into suitable classes, the information on the data sheets will be transferred to working sheets. Further grouping of data and classification according to more than one characteristic may be done wherever possible to make the data more handy, without loss of information. Incidentally, some qualitative information will also be converted into numerical form. A further stage of processing will be the aggregation of data, by working out on computation sheets summary statistics, such as subtotals and totals, proportions, percentages, and indices which stand for broader characteristics of the population or sub-populations. If the data are based on sampling, measures of sampling error will also be computed.

Equipment like punched card machines and electronic computers may be used in large scale work, for carrying out those items of data processing which follow the scrutiny of data. For using punched card machines it is necessary to transfer the information contained in data sheets to cards of a special type, by punching holes at appropriate positions on the cards. Before transferring information to punched cards, information on the data sheets may have to be converted to appropriate codes.

5. *Presentation of results*

This stage of statistical activity deals with the rendering of the results of data processing in a form which will be found suitable for users of statistics. The different types of presentation of statistics are: (1) Tabular presentation, in which statistics are neatly arranged in columns and rows under suitable headings. (2) Graphic presentation, which provides a geometrical representation of numerical information and (3) Textual presentation, in which numerical information is presented together with detailed word descriptions.

The results of special investigations are suitably presented and put together, usually in the form of a report. Statistics compiled as a routine are usually presented in tabular form, sometimes accompanied by charts, and made available as reference tables with brief descriptions about their background. Statistics which are of some importance and which will have many current and future uses are printed and published, unless they are confidential.

6. *Utilization of results*

This phase of statistical activity is concerned with drawing conclusions and making interpretations on the basis of statistical results. Statistics produced and presented by an agency are not always in a form most suited for making decisions and framing policies by another agency. The report on a special investigation, such as a census or sample survey, would contain necessary statistical information together with conclusions and interpretations that will serve the particular purpose for which the investigation was conducted. But even the results of such investigations may sooner or later be utilized for purposes other than those for which they were conducted. Besides, statistics compiled as a routine from administrative records or from published sources, particularly by governments, international agencies and trade associations, are general purpose statistics whose current and future uses are almost unlimited. The treatment such statistics receive will differ according to the purpose for which they have to be utilized. The nature and method of reaching conclusions and undertaking interpretations will to a great extent depend on the type of decisions and actions which are to be based on them. Thus the methods of utilization of statistics for routine administrative control are different from those for planning for development. The main problems in utilization of statistical results are (*a*) selection of appropriate statistics for the purpose specified, (*b*) examination of the statistics as regards adequacy, (*c*) subjecting the statistics to analysis by using efficient methods and (*d*) drawing conclusions and making interpretations which will result in valid decisions being made.

Training on Topic 1.5

7. *Lectures and discussions*

   (1) Presentation of contents of sections 1 to 6 above.

   (2) Familiarization with the details of statistical operations, with respect to selected statistical activities.

8. *Demonstration*

Details of statistical operations. The purpose of this demonstration is to provide an overall picture of various stages and operations involved in a statistical activity. Instead of presenting different aspects of statistical work chosen from several activities, better results can be achieved by confining attention to the different aspects of a single activity although it may not be exhaustive of the variety of statistical work. Even so, it will be necessary to choose at least two activities; one, a special investigation; the other, a compilation of a particular type of routine statistics. The demonstration will accordingly consist of two parts, corresponding to the two types of activity.

Two such activities should be selected and as much material as possible giving information through the different stages of work, obtained in advance. Copies of the material, whenever possible, should be made available to each trainee for reference during the demonstration. They will naturally include the basic records, if any, from which data are extracted, the data sheets, the sampling design, if any, survey calendar, if any, the instruction manuals, field programme, if any, scrutiny programmes, punched card designs, if any, tabulation and other processing programmes, progress charts, presented or published results, and methods, if any, adopted for utilizing the results for specific purposes.

The demonstration, under each part, will consist of an explanation of the various steps taken, starting from the population characteristics aimed at and ending with the final statistical results. Apart from acquainting the trainees with the different details of work, it should also be the purpose of the demonstration to let them know how budgets, data sheets, instruction manuals, punched cards, card designs, code lists, tabulated and charted results, published reports and published statistics look like.

9. *Training material*

Lecture notes. Specimen lists of statistical operations giving details of operations (section 7 above). Demonstration material (section 8 above.)

TOPIC 1.6

Elementary statistical methods—1

1. *Arrangement of data*

The organization of statistical data, as an initial step in the process of evolving statistics out of statistical data, involves some principles. Data, whether large scale or small scale, are made available in an order largely determined by the collection and scrutiny methods. Time series and some other such data may be already in an order suitable for subsequent analysis; but in many other cases, the data will have to be suitably assembled. If they involve only one characteristic, the assembling process is relatively simple. In that case, the data may be ordered according to the magnitude of the characteristic values, or the distinct alternate forms of the characteristic may be ordered and the number of individuals possessing these alternate forms counted and shown against each. Four short data series, each involving only one characteristic, are given below for illustration.

Series I		*Series II*		*Series III*		*Series IV*	
Month	*Power generated (million kWh)*	*City*	*Population (in million)*	*Household*	*Number of members*	*Transaction*	*Type of transaction*
Jan.	43.7	A.	1.7	1.	3	a.	credit
Feb.	42.5	B.	3.2	2.	5	b.	debit
Mar.	47.8	C.	5.6	3.	3	c.	credit
		D.	1.8	4.	7	d.	credit
Apr.	49.6	E.	2.3	5.	2	e.	credit
May	53.5	F.	4.1	6.	5	f.	debit
Jun.	51.7			7.	3	g.	credit
				8.	4	h.	debit
Jul.	50.8			9.	6	i.	credit
Aug.	51.6			10.	3	j.	credit
Sep.	49.3			11.	2	k.	credit
				12.	4		
Oct.	52.6			13.	5		
Nov.	55.4			14.	4		

As regards arrangement of data, the first series may be left as it is, because it is already arranged according to the order of the month. The arrangement of data that may be done on the other three series is shown below.

Series II		*Series III*		*Series IV*	
City	*Population (in million)*	*Number of members*	*Number of households*	*Type of transaction*	*Number of transactions*
C	5.6	2	2	credit	8
F	4.1	3	4	debit	3
B	3.2	4	3		—
E	2.3	5	3	TOTAL:	11
D	1.8	6	1		
A	1.7	7	1		
			—		
		TOTAL:	14		

The arrangement in series II is according to the magnitude of the characteristic values. If there had been two or more characteristics in the data, ordering of the data as a whole according to magnitude could have been done with respect to only any one of the characteristics.

2. *Frequency tables*

The arrangement in the last two cases has resulted in a special type of statistical tables called frequency tables. These tables incidentally achieve some amount of condensation of the data, without losing relevant information. When the characteristic is an attribute, as in the last of the four series above, an additional achievement is quantitizing qualitative data.

When there are two characteristics (attribute and/or variable), two separate one-way frequency tables, or a single two-way frequency table can be prepared. The latter arrangement is more informative. For illustration, a short series of two characteristic data is considered below.

Dwelling unit	*Tenancy status code (owned: 1, rented: 2, others: 3)*	*Number of living rooms*
1	1	3
2	3	4
3	2	3
4	1	5
5	2	2
6	1	3
7	2	4
8	2	2
9	2	3
10	1	4
11	2	2
12	2	3
13	2	2
14	2	3
15	2	2

The corresponding one-way frequency tables (*a*) and (*b*) and two-way frequency table (*c*) are shown below.

(a)

Tenancy status	*Number of dwelling units*
owned	4
rented	10
other	1
TOTAL:	15

(b)

Tenancy status	*Number of living rooms*	*Number of dwelling units*
owned	2	5
rented	3	6
other	4	3
	5	1
TOTAL:		15

(c)

	Number of dwelling units				
Number of living rooms:	*2*	*3*	*4*	*5*	*Total*
Tenancy status					
owned	–	2	1	1	4
rented	5	4	1	–	10
other	–	–	1	–	1
TOTAL:	5	6	3	1	15

In the three arrangements, the information in (*a*) and (*b*) is contained in (*c*); besides, (*c*) provides some further information which can lead to answering questions like: Do dwelling units occupied on a rented basis contain a lesser number of living rooms?

When the data consist of more than two characteristics, it is still possible to have a two-way frequency table, each way being further sub-classified with respect to additional characteristics, or a series of two-way tables arranged in order, so as to result in a multiple two-way classification table. In many types of analysis of data, characteristics are taken out either singly or in pairs for analysis.

3. *Construction of frequency tables*

The preparation of a frequency table essentially consists of (*a*) picking out the distinct alternate forms of the characteristic(s) from the data series and arranging them in some suitable order and (*b*) finding out and entering the frequencies in the appropriate cells governed by the alternate forms.

(1) *Number of classes.* Since one of the purposes in forming a frequency table is condensation of data, the size of the table is important. When the characteristic involved in an attribute, ordinarily there will be only a few alternate forms for the characteristic and all of them will be represented on the frequency table; only in rare cases will a grouping into fewer classes be possible. Also when the characteristic is a discrete variable it may be possible to represent all the distinct values of the variable; but in many situations a grouping into classes will be found necessary to reduce the length of the frequency distribution. If the characteristic is a

continuous variable capable of giving rise to a large number of different values, the grouping the values into a suitable set of classes is almost unavoidable.

No hard-and-fast rule can be given about the number of classes that a frequency table should have. Too few classes may suppress information and too many classes may be unnecessary and lead to undue gaps and irregularities in the frequencies. The regularity with which frequencies will be distributed is one of the determining factors. In general, fewer than 6 or 8 classes and more than 16 classes should rarely be used.

(2) *Selecting class limits:* When the characteristic is a variable, the mid-value of each class is used to represent the class. In choosing class limits, therefore attention may be paid to the fact that the mid-values should, as far as possible, coincide with values around which data tend to be concentrated. For convenience in computation and charting, it is desirable that the class intervals are equal. It may not, however, be practicable to have uniform class intervals, in respect of all data.

(*a*) Discrete variable: If the characteristic in question, is number of passengers arriving at a place by bus per day, and if many of the values show concentration at multiples of 25, the class intervals may be chosen as 38-62, 53-87, 88-112 and so forth with mid-values 50, 75, 100 and so on. If the concentrations are few and far apart, the choice may not be so simple. Approximate and, if necessary, unequal class intervals may then have to be chosen.

(*b*) Continuous variable: In choosing class limits for a continuous variable there are two main considerations: *i*) the continuity of classes will have to be maintained and *ii*) it should be possible to assign each value of the characteristic in the data to its corresponding cell, in a manner consistent with the method adopted for rounding the value. Thus, for instance, if the characteristic is weight and its values appearing in the data have been corrected to the nearest kilogram, resulting in values such as 47, 63, 52, 56 and so forth, and if a class interval of 4 kilograms has been chosen, class limits such as 41-44, 45-48, 49-52 and so on, will be selected having mid-values 42.5, 46.5, 50.5 and so on. While these class limits will create no difficulty in the construction of the frequency table from the data, for the sake of clarity about the continuous nature of the variable, the class limits may be instead written as 40.5-44.5, 44.5-48.5, 48.5-52.5 and so on, making note of the fact that overlapping of class limits does not mean overlapping of values in the data.

When, however, the values of the variable are rounded in a different way, suitable adjustments are necessary. Thus for example, if the data consist of values of age *last birthday* and not age nearest birthday, and the class interval is 10 years, suitable class limits will be 0 and under 10, 10 and under 20, 20 and under 30 and so on.

4. *Graphic representation*

Statistical results may be charted for the purpose of final presentation or as an aid to further analysis of the results. Some of the graphic representations associated with frequency distributions are the bar diagram, the histogram, frequency polygon, frequency curve, ogive and the scatter diagram. These are described in Topic 6.2.

Training on Topic 1.6

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Presentation of different types of one-way and two-way frequency tables.

(3) Illustration of construction of frequency tables, with emphasis on choice of class intervals.

(4) Illustration of construction of simple charts associated with frequency tables, and some of the uses of these charts.

6. *Exercises*

Examples:

(1) (*a*) Preparation, by frequency plotting (tally marks), one-way frequency tables from given single-characteristic series, one each for: (*a*) attribute, (*b*) count and (*c*) measurement. Each series given may involve about 100 individuals. In the case of measurement, the series will be such as to require grouping of the characteristic into class intervals.

(*b*) Calculation of percentages from the frequency table based on attributes.

(*c*) For the measurement data, plotting the histogram and ogive, and using the ogive to find the median and other fractiles.

(2) (*a*) Preparation, of two-way frequency tables from multiple characteristic series. A data series will be given relating to about 50 individuals on each of which three characteristics, (one attribute and two variables) have been observed. Four separate two-way frequency tables will be prepared by frequency plotting, the first three corresponding to the three combinations of the characteristics taken two at a time and the fourth representing all the three characteristics in a single two-way multiple classification table (i.e. along one of the two ways, two of the characteristics will be shown by sub-classification).

(*b*) Plotting the scatter diagram for two of the characteristics from the original series.

7. *Training material*

Lecture notes. Exercise sheets. Specimen frequency tables and charts (section 5 above).

TOPIC 1.7

Elementary statistical methods—2

1. The aggregation of characteristics observed on individuals will be performed with respect to all the individuals either in a finite population or in a sample. The result of such aggregation may be a total of counts both when the characteristic observed on the individual is an attribute or a count. The result may be a total of measurement values when the individual characteristic is a non-count variable. All these totals will be meaningful as population characteristics when the aggregation is over a finite population. But when the aggregation is over a sample, the sample characteristics are more meaningfully expressed on the basis of 'standards of reference' such as a per-individual, per-hundred or per-thousand individuals basis. Accordingly, a sample characteristic derived from attributes will be expressed as a proportion, percentage or rate; a sample characteristic derived from counts or non-count variable values will be expressed as an average, or as some other per-individual value.

2. *Univariate data*

In the analysis of data relating to one variable, some of the characteristics that are commonly used are: 1) mean, 2) median, 3) mode, 4) range, and 5) standard deviation. These may be either sample characteristics or population characteristics. The square of the standard deviation is called variance, especially when it is a population characteristic.

The mean, median and mode are characteristics either expressed as a per-individual value or chosen as a typical value of the individual characteristic.

The *mean* of a set of values is the arithmetic average. Thus the mean of the values, 3, 4, 8, 9, 8, 4, 2, 3 is $\frac{41}{8} = 5.125$.

The *median* of a set of values is the middle value obtained after the set of values have been arranged in ascending or descending order of magnitude. Thus the median of the values 4, 7, 4, 3, 6, 5, 4, 5, 9 is 5, obtained after arranging the values as follows: 3, 4, 4, 4, 5, 5, 6, 7, 9, and the median of the values 4, 7, 2, 3, 5, 2, 6 is $\frac{5+4}{2} = 4.5$, obtained after arranging the values as follows: 2, 2, 3, 4, 5, 5, 6, 7.

The *mode* in a set of values is that value which occurs most often. Thus the mode among the values 5, 4, 12, 7, 5, 6, 5, 12, 10, 4, 5, 3 is found to be 5.

The range and standard deviation do not typify the individual characteristic, but are characteristics standing for the variation between individual characteristic values.

The *range* of a set of values is the difference between the largest value and the smallest value in the set. Thus the range of the values 12, 7, 5, 3, 2, 15, 6, 8 is $15 - 2 = 13$.

The *standard deviation* of a set of values is obtained as follows: the mean of the values is first calculated; the difference (deviation) from the mean, of each value is taken; each such deviation is squared; the mean of the squares is then found out; the square root of this mean is the standard deviation. (The standard deviation is sometimes called the "root mean square deviation", which is indicative of the method of calculation). Thus the standard deviation of the values 5, 3, 4, 6, 5, 2, 3 (of which the mean is $\frac{28}{7} = 4$) is

$$\sqrt{\frac{1^2 + (-1)^2 + 0^2 + 2^2 + 1^2 + (-2)^2 + (-1)^2}{7}} = \sqrt{\frac{12}{7}} = 1.309.$$

The nature of information conveyed by characteristics like the range and standard deviation can be seen from the following. The two sets of values 7, 10, 8, 11 and 9, 10, 9, 8 have the same mean, namely 9. (In this example, they have the same median also which again happens to be equal to the mean). But the variation among the values (dispersion of the values) is different in the two sets. The variability as measured by the range is 4 for the first set and 2 for the second; and as measured by the standard deviation 1.58 and 0.71 respectively.

The weighted mean: In calculating the mean of the values 5, 6, 5, 2, 6, 5, 5, 6, instead of adding them and dividing by 8, the required mean can be obtained as the mean of 5, 6 and 2 each weighted by its respective frequency of repetition, namely, 4, 3 and 1, as follows:

$$\frac{4 \times 5 + 3 \times 6 + 1 \times 2}{4 + 3 + 1} = \frac{40}{8} = 5.$$

The mean of the 5 values 3, 3, 5, 2, 7 is 4 and the mean of the 6 values 2, 4, 3, 3, 5, 1, is 3. From these two means, the mean of the combined set of 11 values can be obtained as the *weighted* mean of the two means as follows:

$$\frac{5 \times 4 + 6 \times 3}{5 + 6} = \frac{38}{11} = 3.45.$$

In a group of 150 people, the proportion of males is 0.48 and in another group of 250 people the proportion of males is 0.51. The proportion of males in the combined group of 400 people is given by the weighted mean of the two proportions, as follows:

$$\frac{150 \times 0.48 + 250 \times 0.51}{150 + 250} = 0.50.$$

3. *Index numbers*

Index numbers are figures, generally expressed as percentages, calculated for purposes of comparison of similar magnitudes, such as the price level over different periods of time or different localities.

If the production of a commodity was 53 thousand units in January and 60 thousand units in February and 61 thousand units in March, the ratios $\frac{60}{53}$ and $\frac{61}{53}$ are index numbers of production for February and March with January as the base for comparison. In other words, if the index for January is 100, the index for February is $\frac{60}{53} \times 100 = 113$ and the index for March is $\frac{61}{53} \times 100 = 115$. In his example, the magnitudes relating to one commodity or item are only compared. In many situations, the magnitudes for comparison may be composite and based on magnitudes relating to many commodities or items.

Example:

Price in dollars

Year	Rice (per kilogram)	Cloth (per metre)	House rent (monthly)
1950	0.14	0.86	75
1962	0.12	0.78	83

Monthly family expenditure may be considered as the magnitude for comparison and the three items may be considered as its main components. In such a case, one of many alternative procedures may be adopted. Some of the alternatives are shown below:

(1) Index number as ratio of simple totals:

The simple totals are:

1950: 76.00
1962: 83.90

Index numbers (1950 as base):

1950: 100

$$1962: \frac{83.90}{76.00} \times 100 = 110$$

(2) Index number as simple arithmetic mean of ratios (called relatives in this context).

The respective relatives are:

Rice

1950: 100

$$1962: \frac{0.12}{0.14} \times 100 = 85.7$$

Cloth

1950: 100

$$1962: \frac{0.78}{0.86} \times 100 = 90.7$$

House rent

1950: 100

$$1962: \frac{83}{75} \times 100 = 110.7$$

Index numbers (1950 as base):

1950: 100

$$1962: \frac{85.7 + 90.7 + 110.7}{3} = 96$$

(Instead of the arithmetic mean of relatives, it is also the practice to use the geometric mean which has some good properties in respect of index numbers. The geometric mean of n values is the nth root of the product of the n values).

The index numbers in the two cases above have the weakness that the relative weights with which the three items enter into the main magnitude are not taken into account. Suppose that 26.5 kilograms of rice and 7.2 metres of cloth are the quantities consumed in a month per family. For the sake of simplicity, let these values be applicable for 1950 as well as 1962. Then, weighted index numbers may be constructed.

(3) Index numbers as ratios of weighted totals:

The weighted totals are:

$$1950: \quad 26.5 \times 0.14 + 7.2 \times 0.86 + 1 \times 75 = 84.90$$

$$1962: \quad 26.5 \times 0.12 + 7.2 \times 0.78 + 1 \times 83 = 91.80$$

Index numbers (1950 as base):

1950: 100

$$1962: \frac{91.80}{84.90} \times 100 = 108$$

(4) Index number as weighted arithmetic mean of relatives:

Supposing that the 1962 values (i.e. price × quantity) are used as weights in this case, the values are:

Rice	*Cloth*	*House rent*
26.5×0.12	7.2×0.78	1×83

Index numbers (1950 as base):

1950: 100

$$1962: \frac{26.5 \times 0.12 \times 85.7 + 7.2 \times 0.78 \times 90.7 + 1 \times 83 \times 110.7}{26.5 \times 0.12 + 7.2 \times 0.78 + 1 \times 83} = \frac{9969.99}{91.80} = 109.$$

(Instead of the weighted arithmetic mean of relatives, it is also the practice to use the weighted geometric mean. In that case, weights will be applied to the logarithms of the relatives.)

4. *Mean and standard deviation from a frequency distribution*

Calculation of the mean and standard deviation of a variable characteristic may be done from a frequency table, if the data have already been arranged in one. These are obtained by calculating the mean and standard deviation of the alternate forms of the variable as represented on the frequency table, making provision for the frequency of repetition of the alternate forms. In other words, in obtaining the means (of the values or squares of the values, of the variable) the weighted means should be taken, the weights being the respective frequencies. It is easily seen that weighted means obtained by using the frequencies as weights is the same as weighted totals using the relative frequencies as weights. Thinking in terms of relative frequencies is more appropriate when the means and standard deviations refer to the population. In calculating the mean and standard deviation from a frequency table with class intervals, the mid-points of the classes will represent the alternate forms.

5. *Moving averages*

The method of moving averages consists in obtaining from a given series of values a new series in which each value is the mean of a number (always the same number) of values in the original series. The method is particularly useful in smoothing out short term fluctuations in time series. The calculation of a 5-point moving average series is shown below.

Month	*Percentage of workers unemployed*	*Sum of five values*	*Moving average*
April	2.5	...	...
May	3.3	...	...
June	4.0	...	4.1
July	4.7	...	4.6
August	6.0	20.5	4.7
September	5.0	23.0	4.6
October	3.6	23.3	5.2
November	3.7	23.0	5.6
December	7.8	26.1	5.5
January	7.7	27.8	5.4
February	4.7	27.5	5.3
March	3.0	26.9	4.1
April	3.2	26.4	3.3
May	2.1	20.7	2.5
June	3.3	16.3	...
July	1.1	12.7	...

Training on Topic 1.7

6. *Lectures and discussions*

(1) Presentation of contents of section 1 to 5 above.

(2) Illustration of calculation of summary figures from data. (Symbols and formulae and use of calculating machines may be avoided under this topic. They will be introduced in Topic 5.2).

(3) Illustration of calculation of simple index numbers and moving averages from data (without use of formulae and machines).

7. *Exercises*

Examples:

(1) Calculation of mean, mode, median, range and standard deviation from univariate data in ungrouped form. Two different sets will be calculated from two separate series each consisting of about 20 or 15 observations. (Hint to be given in simplifying calculations by changing the origin and/or scale of the variable values).

(2) (*a*) Calculation of mean and standard deviation from given frequency table. The calculation will be done twice, first by using frequencies as weights and next by calculating relative frequencies and using them as weights (Hint to be given regarding change of origin and scale).

(*b*) Calculation of simple index numbers.

(*c*) Calculation of moving averages of specified length, from a given series.

8. *Training material*

Lecture notes. Exercise sheets.

TOPIC 1.8

Elementary statistical methods—3

(for intermediate level only)

1. *Probability*

The role of probability theory in statistics, is in relation to sampling. If a figure is chosen at random from the figures 1 to 100, the probability of the figure having 7 as its last digit is $\frac{1}{10}$. The probability is known in this case, because the composition of the set of items from which the item is chosen at random is known. Knowing the probability of a result to be $\frac{1}{10}$ does not, however, mean that if 10 attempts are made, the result will definitely occur once. Thus, whereas there will be exactly 10 figures ending in 7 if all the figures 1 to 100 are taken together, if a hundred successive random selections are made from the set of figures 1 to 100, the number of figures ending in 7 may not be exactly 10; nor, if this is done two hundred times, will the number surely be exactly 20. But as the number of repetitions increases, the ratio $\frac{1}{10}$ will be 'approached more and more closely'. Probabilities are the best available means for making statements about results surrounded by uncertainty; and their validity increases with increase in the number of 'exposures' for the result to occur.

Probability is expressed as a numerical value in the range 0 to 1, 0 standing for complete uncertainty and 1 for complete certainty. The probabilities of a full set of exclusive alternate results add up to 1. There are some rules useful in calculating probabilities, knowledge of which will also be required in dealing with samples. The first is the *addition rule* of probability which states that, if there are some alternate results which are mutually exclusive (i.e. they cannot happen together), the probability of any one of the results occurring, is the sum of the separate probabilities of the results. Thus, if a figure is being chosen at random from the figures 1 to 100, the probability of a figure which is even (i.e. ending in 0, 2, 4, 6, or 8) is $\frac{1}{10}+\frac{1}{10}+\frac{1}{10}+\frac{1}{10}+\frac{1}{10}=\frac{1}{2}$ or the probability of a figure which is either even or ending in 5, is $\frac{1}{2}+\frac{1}{10}=\frac{3}{5}$. But it will not be correct to say that the chance of choosing either an even number or a multiple of 5 is $\frac{1}{2}+\frac{1}{5}$, because the two results overlap (i.e. they are not mutually exclusive). The probability of this result will be $\frac{1}{2}+\frac{1}{5}-\frac{1}{10}$ where $\frac{1}{10}$ is the probability of an even number being also a multiple of 5.

The second rule is the *multiplication rule* which applies to results occurring independently. Results are said to be independent if the occurrence of any one among the results does not alter the *probability* of occurrence of the other results. The multiplication rule states that the probability of a specific result repeating independently or different results occurring independently in a particular sequence, is the product of the probabilities of separate occurrences. Thus, if the probability of a man aged 55 years surviving 10 years is $\frac{3}{5}$ and of a woman aged 50 years surviving 10 years is $\frac{5}{9}$, the probability that both will survive 10 years is $\frac{3}{5} \times \frac{5}{9} = \frac{1}{3}$. If three figures are selected at random, each time independently, from the figures 1 to 100, the probability that all the three figures end in 9 is $\frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{1000}$. If three figures are selected at random, each time independently, from the figures 1 to 100, the probability that the first figure ends in 3, the second in 7 and the third in 8 is $\frac{1}{10} \times \frac{1}{10} \times \frac{1}{10} = \frac{1}{1000}$. The multiplication rule is easily modified to suit dependent results, by considering *conditional probabilities.* The multiplication rule for dependent events states that the probability of both of two results occurring is the unconditional probability of the first result multiplied by the conditional probability of the second. If, in the random selection from figures 1 to 100, the figure chosen in the first selection is excluded in the second selection, then the results of the two selections are not independent. The probability of obtaining figures ending in 6 in both the selections, is given by $\frac{1}{10} \times \frac{9}{99} = \frac{1}{110}$ and the probability of obtaining a figure ending in 6 in the first selection and a figure ending in 7 in the second selection is $\frac{1}{10} \times \frac{10}{99} = \frac{1}{99}$. The multiplication rule for dependent events may also be used to find the value of conditional probability. Thus in the random selection from figures 1 to 100, the probability of obtaining a multiple of 5, under the condition of its being even, is given by $\frac{10}{100}$ divided by $\frac{50}{100} = \frac{1}{5}$.

2. *Population and sample*

The individuals in any population, finite or infinite, will fall into different exclusive classes corresponding to the different alternate forms which the individual characteristic can take. Accordingly, there will be a distinct proportion (relative frequency) of individuals in the population, associated with each class. Such a proportion associated with a particular class is itself the probability that an individual selected from the population will belong to that class, if during selection each individual had the same chance of being selected. The above probabilities will serve as a means to determine population characteristics from individual characteristics. Thus, with respect to the attribute employment, the individuals may fall into the three classes: fully employed, partially employed and unemployed. The population characteristics in this case will be the proportion of persons in these classes. With respect to the variable household size, the different classes will be determined by the different values of the variable, namely, 1, 2, 3, . . . up to a certain number. Two of the useful population characteristics in this case will be the mean and standard deviation of the variable obtainable on the basis of the proportions (the relative frequencies) of individuals in the different classes. The same applies to a continuous variable like income, in which case the different classes will be determined by suitable class intervals.

The above characteristics will not be known for the population, unless the characteristic is observed on all the individuals in the population (possible only in a finite population). A random sample consisting of observations of the characteristic on a number of individuals, leads to a frequency table and provides estimates of the relative frequencies in the population and therefore estimates of the population characteristics. In the case of an individual characteristic which is a variable, the population characteristics may also be estimated from the sample characteristics calculated without forming a frequency table.

3. *Standard errors*

In using a sample characteristic to estimate a population characteristic, sampling error is unavoidable. This can be seen from the fact that estimates based on two different samples from the same population may not be identical. It is therefore necessary to give, along with a sample estimate, an indication of the sampling error. It is not possible to give the exact value of the error because the population characteristic is not known. However, it is possible to give a statistical 'measure' of this error. One such measure of sampling error associated with a sample characteristic is the standard error of the sample characteristic. The population characteristic can be expected, with very high probability, to be with a margin of twice the standard error on either side of the sample characteristic. The magnitude of the standard error will depend on the sample size, and can be reduced by increasing the sample size. The formulae for calculating the standard errors of the sample mean and the sample proportion are $\sqrt{\frac{s}{n}}$ and $\sqrt{\frac{p\ (1-p)}{n}}$ respectively where s and p are estimates of the standard deviation and proportion in the population respectively. These formulae are applicable to the situation when the sample is an unrestricted random sample (equal probability selection) either from an infinite population, or from a finite population with replacement of the individual selected each time. When the sample is from a finite population of size N and selected with replacement, each of the above formulae may be multiplied by the factor $\sqrt{\frac{N-n}{N-1}}$.

Training on Topic 1.8

4. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 3 above.

(2) Consideration of simple problems requiring the use of addition and multiplication rules of probability.

(3) Introducing the concept of a probability distribution of a variable and uses of the concept.

(4) Introducing the concept of a sampling distribution and uses of the concept.

5. *Demonstrations*

Examples:

(1) *The meaning of probability.* This demonstration is intended to clarify ideas about probability of a result and show its importance in sampling by simulation of sample results through coin tossing, die throwing and playing card selection experiments. The face turning up when a coin is tossed, the figure appearing on top when a die is thrown, or the symbol or number on a card selected from a shuffled pack of cards is a characteristic which can take alternate forms. Each toss, throw, or selection is equivalent to selecting an individual at random from a population. Tosses, throws or selections would thus constitute random samples.

In experiments with the coin, die or cards, the probability of a particular result (equivalent to a specific alternate form of the characteristic) is known beforehand, at least approximately, if not precisely due to the coin or die being biased. Experiments will be conducted during the demonstration to illustrate the following:

(*a*) Probabilities are 'approached more and more closely' with more and more trials (or selections).

(*b*) The meaning of dependent and independent results of trials.

(*c*) The meanings of the results to which the addition and multiplication rules of probability, including conditional probability, are applicable.

(*d*) It is possible to visualize a new population constituted by the totality of samples, (sampling distribution). This population may be studied in order to judge the reliability of a sample characteristic obtained in a sample.

(2) *The meaning of standard error.* This demonstration will consist of two parts (*a*) presentation of frequency distributions corresponding to sampling distribution and (*b*) verification of formulae for standard error of the mean.

For the first part, a number of samples (at least 50 samples each of size 16) should be drawn, using the apparatus described in the demonstration under Topic 1.4. These samples will give rise to 50 sample characteristics of each of the three types (i.e. 1. sample proportion, 2. average number of dots and 3. average of value). From these, three frequency tables should be prepared. The results should be made ready before the demonstration.

The three frequency tables will be presented at the demonstration and how they were prepared will be explained. The fact that these frequency tables are tables relating to a sample of sample characteristics and that they correspond to the sampling distribution (i.e. the relative frequency distribution representing the population of samples and not the population of the original individuals) should be clearly explained. Further, it should be made clear that the standard deviation of the sample characteristic in the population of samples can be used to measure the reliability of a sample characteristic calculated from a sample. This standard deviation cannot be numerically evaluated unless the population of original individuals, or the population of samples is completely observed; and so it will have to be estimated. (This standard deviation is sometimes referred to as the standard error of the sample characteristic. But it is more correct to call its estimate standard error).

The second part of the demonstration will be to verify the formulae for standard error of the mean with the help of a very small finite population, and exhausting all samples, so that the sampling distribution will be actually known. This is done by taking a finite population of size 4 and samples of size 2 from it.

The characteristic values of the 4 individuals in the population may, for example, be taken as 2, 3, 5, 2.)

All the 6 samples corresponding to sampling without replacement and the 16 samples corresponding to sampling with replacement will be enumerated. In each of the two cases, the sampling distribution of the mean will be prepared, and the standard deviation in that distribution calculated and compared with that given by the respective formula (i.e. the formula for standard error of the mean, in which the value of s will be taken as the standard deviation in the population of original individuals and not the sample estimate). (The necessary computations may be got ready before the demonstration is made).

6. *Exercises*

Examples:

(1) (*a*) Working out simple probability problems utilizing the addition and multiplication rules of probability, including conditional probability.

(*b*) Calculation of population mean and standard deviation (expectation and variance) from simple numerically expressed probability distributions.

(2) Enumerating all possible samples (equally likely cases) of size 2 from a finite population of size 5, with respect to one variable characteristic, corresponding to the two cases: (*a*) sampling without replacement (10 samples possible in all) and (*b*) sampling with replacement (25 samples possible in all) and verification of the formula for standard deviation of the mean applicable to each of the two cases.

(Only simple values of characteristics in the population should be given, such as 4, 3, 2, 3, 5.)

7. *Training material*

Lecture notes. Exercise sheets. Demonstration material (section 5 above).

TOPIC 1.9

Qualities of good statistics

1. A number of decisions are made and a series of operations gone through before statistics of any type are finally obtained. There are thus several possibilities of errors being made which can affect the quality of statistics and invalidate them partially or fully. Precautions and controls are, for this reason, exercised at all stages of statistical work. In spite of this there may be still some limitations beyond control. Such limitations, therefore, go with the statistics and are pointed out, whenever known, for the guidance of users.

Some of the criteria by which the quality of a given set of statistics can be judged are: (*a*) the accuracy of the figures, (*b*) validity of different characteristics as measures of the concepts they are intended to measure, (*c*) uniformity with respect to time reference or other qualifications of the characteristics, (*d*) completeness with respect to coverage of the entire population or sample.

The different types of errors over which control may be exercised are: errors due to faulty concepts and definitions, errors due to faulty selection of individuals, observational errors, processing errors, presentation and publication errors and errors due to faulty overall designing.

2. *Concepts and definitions*

Standard concepts and definitions are necessary not only to ensure adequacy and uniformity within the same set of statistics but also for comparability between different sets of statistics of the same type. It is necessary to settle beforehand questions like: whether personal income should include income from all sources or only salaries and wages; whether an unemployed person means only one who has registered with the employment exchange or not; whether a hotel or military barracks is to be treated as a dwelling unit or not; whether a concept like health should be measured by the characteristics death rate and disease rate only or by some others also; whether investment statistics should relate to the business year or the calendar year.

3. *Selection of individuals*

The coverage with respect to the individuals should be complete, free from omissions, repetitions, and inclusion of extraneous individuals. For example, omissions and unnecessary inclusions can take place when a list of individuals which is not up to date, is used. Repetitions of families can occur when families are selected with the help of a list of wage earners. Sometimes a whole class of individuals may get omitted, as for instance, in giving statistics of growth of trade only statistics of growth of export may be included and not those of internal trade.

When sampling is used, the population will be properly covered only if the sample is a representative one. A biased sample will lead to misleading statistics. Examples:

(*i*) After a census of human population is made, the data sheets relating to the members of each family are separately bundled with that of the head of the family on top. If from such bundles a sample of sheets is selected by picking out every fifth sheet starting from the top, it will be found that the sample contains a very high proportion of sheets relating to males, because the head of the family is usually a male and the size of a family is often close to five.

(*ii*) A sample that is representative of a population with respect to the characteristic 'economic status' may not be representative with respect to the characteristic 'religious affiliation'.

A sample too small in size may not also give rise to reliable statistics.

4. *Method of observation*

The most efficient method of observation that is possible in a given situation should be used, in order to ensure good quality statistics; measurements and counting are efficient methods; when they are not possible, interviews and personal judgment may have to be resorted to. In making measurements, errors can occur in several ways such as by: using non-standard instruments, failure to define and conform to units of measurement, not using the full accuracy attainable with the instruments given, introducing personal biases in making reading and rounding off values. Answers based on interviews can be seriously affected by factors such as: who asks the question, or how it is asked, or for what purpose it is asked. Considerable accuracy in personal judgment can be attained with skilled staff.

5. *Processing*

Adequate checks should be provided for controlling processing errors such as errors in copying and in computation. In some computational procedures built-in checks may be provided. But in general, at least an independent sample check will be necessary. In some cases, it will be worth while to duplicate the processing work independently. When punched card machines are used, the punching is verified often completely and suitable checks are introduced in subsequent work.

Processing errors can occur also due to lack of appreciation of analytic methods, as for instance in not providing for the weights in computing an average of averages or an average of percentages, based on unequal numbers.

6. *Presentation and publication*

Statistics should be presented together with their full background, indicating all known limitations and care exercised in avoiding inadequate or misleading methods of presentation and fallacious interpretation of results, besides copying and printing errors. Presentation of percentages, or numbers without the appropriate 'standards of reference', failure to give totals wherever possible, use of abnormal scales in charts meant for comparison, and interpretations based on spurious association are some of the features which can reduce the value of the statistics.

7. *Designing*

Faulty overall designing of statistical work can also result in reducing the value of statistics considerably, as for example, if too long an interval elapses between data collection and publication, or if the statistics made available do not justify the cost incurred to obtain them.

Training on Topic 1.9

8. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 7 above.

(2) Illustration of different factors affecting quality of statistics, with the help of selected examples.

9. *Demonstration*

Example:

The implications of measurement. This demonstration would bring out the nature of inaccuracies that can influence measurements and point out that increasing the precision of the measuring instrument and taking precautions are essential for accuracy.

The apparatus required for this purpose consists of a large board painted in light colour and with black grid lines, drawn on its upper half portion, as on graph paper, and a few wooden blocks, different in length and thickness. There should be some arrangement such as a pair of sliding tight rubber bands to fix one block at a time in any position on the face of the board. One of the blocks is fixed in the middle horizontal line on the board such that only the edges away from the vision of the trainees are in contact with the grid lines. The trainees (sitting in different positions in the class facing the board) are asked to read and record the length of the block in terms of the unit of the grid division, as accurately as possible. The reading is repeated at least once and the mean is taken. The mean measurement obtained by each trainee is posted in tabular form (on the blackboard), broadly classified according to the seating position of the trainees. The process is repeated with another block which is different in length and a third one which is different in thickness.

The results are compared. Considerable variations in the measurements will be noticed. The variation will be due to a number of reasons, such as the seating position of the trainees, the unit of measurement, the thickness and length of the blocks, errors in reading, method of rounding, other personal errors and so on. The similarity of measurement to other forms of observation and the need for precautions in making observation, as also the need for tolerating some unavoidable errors should be pointed out.

10. *Training material*

Lecture notes. Demonstration material (section 9 above).

Course 2

DESIGNING A STATISTICAL INVESTIGATION

This course is intended for both the primary and intermediate levels. The main purpose of the course is to provide a general acquaintance with some principles of designing and to provide knowledge of the detailed items of work involved in designing an investigation.

Although the subject of designing a routine compilation project would form part of Course 4, and the present course is entirely on designing special investigations, training gained in this course will have many aspects that will be ueful in compilation work.

Training will be given through lectures, discussions, demonstrations and exercises.

Course summary:

Topic	*Number of hours proposed for instruction*			
	Lecture/ Discussion	*Demonstration*	*Exercise*	*Total*
2.1 Principles of designing	6	4	4	14
2.2 Details of designing	8	2	2	12
TOTAL:	14	6	6	26

TOPIC 2.1

Principles of designing

1. Designing is necessary to ensure that the objectives of the statistical activity are met with the optimum use of resources. Besides overall designing, the different stages of work, such as data collection, data processing and presentation and publication of results, require designing. There is, however, some difference between designing a statistical investigation and designing statistical activity concerned with routine compilation. In the former, there is scope for full designing with a view to producing statistics that will serve certain specified purposes; while, in the latter, complete designing may not be possible as, for example, the nature and coverage of information may have already been determined by the records and other material for compilation.

2. *Steps in designing*

The total resources available for conducting the investigation being known, the first step in designing is to visualize the purposes which the statistics, when obtained, will have to serve. It will then be possible to resolve the requirements into items of statistical information expressible as population characteristics. The following alternatives are then examined as means to obtain these characteristics: (1) abstracting statistics already available in some form or other, (2) compiling from existing administrative records, (3) collecting data first hand by calling for statistical returns, by mail or otherwise and (4) collecting data first hand by field work. The following considerations are then carefully made: (*a*) the workability of each alternative, (*b*) the characteristics in terms of which information will be required and entities to which these characteristics would relate, under each alternative, (*c*) whether information on a sampling basis would suffice or the entire population is to be covered, (*d*) the reliability of results that each alternative will lead to, (*e*) the time and cost involved in each alternative. On the basis of these considerations and keeping the resources available in view, a choice of one of the alternatives or of a combination is made. The subsequent stages of designing are as follows. Concepts and definitions relating to the characteristics and individuals are settled and suitable data sheets prepared. Procedures for data collection, data processing, presentation and utilization of results and publication, are laid down. Programme and organization of each stage of work are worked out and arrangement for the administration of the investigation made. Instruction manuals relating to the different operations in investigation are prepared.

3. *Sampling*

Sampling is a method of reducing the cost of an investigation without proportionate loss in the accuracy of results. In large-scale work, therefore, sample surveys may be conducted instead of complete censuses. The presence of sampling error in results based on sampling will not be a serious factor in many cases; because, firstly, the sampling error can be controlled within reasonable limits and secondly, even in a complete census, some errors called non-sampling errors will exist. Non-sampling errors occur due to various reasons and are difficult to control in a complete census especially when the population is very large. Non-sampling errors can occur also in sampling, but they can be more effectively controlled in a sample. Accordingly, one of the important steps in designing an investigation is to weigh the consequences of sampling against those in a complete census and to decide which method will ultimately be more profitable.

For designing an investigation on a sampling basis, some prior information about the characteristics is necessary. If this is not available, a pilot investigation on a small scale will have to be conducted.

Basic concepts about sampling were given under Topic 1.4 and some further ideas under Topic 1.8. For the theory of sampling to hold good, it is necessary that the selection of individuals constituting the sample is random. (Random selection is otherwise known as probability selection. Selection which is not random is called purposive selection). The element of randomness will be present in most types of statistical samples. There are, however, many variations of sampling, each having its own features and suitability in a given situation. It is the procedure of selecting the individuals that often determines the name of the sample type. There are thus simple random sampling, cluster sampling, multistage sampling, stratified sampling, systematic sampling and several other types (Demonstration 1 below). Methods and formulae for calculating sample characteristics from sample data will differ according to the type of sampling; so also the calculation of measures of sampling error associated with the sample characteristics will differ.

4. *Designing data sheet*

In designing a data sheet, several principles are involved. Some of the requisites of a good schedule or questionnaire are the following: (1) It should be as short as possible. (2) Each topic and item or question should be arranged in proper sequence with attention to the logical and psychological flow of thought involved in answering questions. (3) Each item or question should be clear allowing only a single interpretation. (4) Items and questions should be capable of being easily interpreted. (5) Questions should not be asked in a manner that will influence the answers. (6) Offending questions should be avoided. (7) Sufficient space should be allowed on the data sheet for answers and subsequent totalling and other calculation and for coding. (8) The lay-out of the data sheet should be such as to facilitate work like sorting and tabulating. (9) Built-in checks should be provided where possible.

Designing the data sheet will have to be based on a list of characteristics for observation, giving clear descriptions of these characteristics, the population and the individuals laid down in conformity with selected concepts and definitions. Careful consideration will have to be given, depending on the methods of data collection and observation to be adopted, as to the arrangement of entry headings or questions and the manner in which answers to questions and other entries will be made.

5. *Cost*

Since one of the chief aims in designing is efficient utilization of resources for obtaining statistical results, considerations of the costs involved in the different operations and the overall cost are of great importance. Thus, in designing it will be necessary to choose between alternate procedures for each stage of the investigation, after carefully weighing the possible achievements of the procedures against the cost to be incurred. For example, in a complete census, some amount of error in the results may be unavoidable due to non-sampling errors and the cost of conducting the census may be large. In such a situation, it may be possible to conduct a sample survey at a much lower cost which will produce results within an allowable margin of error. As another example, since punched card methods involve the use of different types of machines at high rental or price, it will not be economical to adopt them for data processing unless the investigation is on a large scale and therefore capable of continuously feeding the different machines.

Training on Topic 2.1

6. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 5 above.

(2) Familiarization, in a general way, with sampling schemes adopted in selected investigations.

(3) Consideration of specimen data sheets, from the point of view of details of their construction.

7. *Demonstrations*

Examples:

(1) *Sample types and selection procedures.* The purpose of the demonstration will be only to show what each type of sampling means in respect of the procedure for selecting the individuals.

The apparatus required is a tray containing 100 beads of equal size of which 20 are of one colour (pink), 30 of another colour (blue) and the other 50 of a third colour (white). Ten small boxes each capable of holding 10 or more beads are also required. Different types of sampling will be shown assuming that in each case a sample of a predetermined size will be chosen from the population of 100 beads. It may be assumed that the characteristic of interest on each bead is its weight. Since the beads are not labelled with serial numbers, it will not be convenient to identify beads while in the tray, with random numbers selected from a table of random numbers. One convenient way of selecting a simple random sample in such a situation is first to select a suitable set of random numbers, arrange them in serial order, then take out each unit without paying attention to any property of the unit, one by one out of the tray, counting serially and keeping aside those units which correspond to the selected random numbers. This method of random selection may be adopted throughout the demonstration. Sampling with replacement or without replacement may be adopted.

1) To select a simple random sample of size 20 from the population of 100, twenty two-digited random numbers will be copied out from random number tables and arranged in ascending order. (00 being considered as 100). The 100 beads will be mixed in the tray and counted out one by one; those corresponding to the random numbers are then kept separately from the rest.

2) For selecting a systematic sample of size 5, with a random start, one random number which is 20 or less will be first selected. Let it be 15. Then, while the counting out is being done, the 15th, 35th, 55th, 75th and 95th beads will be kept separately to form the required sample.

3) To select a stratified sample, the stratification being according to colour, of size 10, consisting of 3 beads from the pink, 3 from the blue and 4 from the white stratum, the beads will be grouped according to colour and each colour group will be treated as a sub-population from which sub-samples of sizes 3, 3 and 4 respectively will be chosen at random.

4) To select cluster samples to give a total sample of size 20, the beads will be mixed and 10 beads will be put into each of the 10 small boxes. Two boxes will be selected at random, by counting out the boxes this time. The 20 beads in the two boxes will form the sample.

5) To select a two-stage sample, (in which 2 first stage units and from each of those two units 5 second stage units are required, making a total sample size 10), the beads will be mixed and distributed in the 10 boxes with none containing less than 5 units. Two boxes will be first selected at random and from within each of the two boxes, 5 beads will be similarly selected at random.

(2) *Sample versus complete census.* Considerations of merits and demerits of sampling as an alternative to a complete census can be reduced to the following. Sampling involves less cost, a lesser degree of non-sampling error, but some amount of sampling error is unavoidable. A complete census is much costlier, is free of sampling error, but not free from non-sampling error. This situation allows the possibility that in sampling, results subject to some margin of error can be obtained at relatively low cost. Considering the total error in the results and the cost to be incurred together, it is possible that sampling is, on the whole, more advantageous than complete census. It is also not impossible in certain cases, depending on the nature of the non-sampling error that sampling involves less cost and at the same time more accuracy than complete census. These can be demonstrated by developing a tabular comparison of the type shown below. The following, which are taken for granted in the table, should be first explained.

(1) The errors in question may be expressed as relative (percentage) errors.

(2) Sampling error decreases as sample size increases.

(3) Non-sampling error is negligible in very small samples and increases as sample size increases.

(4) The total error may be taken as the sum of sampling error and non-sampling error.

In the table below, the cost of the investigation is assumed to be a constant multiple 0.5 of the number of individuals observed and the population size is taken to be 25,000. (The figures used for the non-sampling error values in the table are hypothetical and used only for the purpose of illustration of possibilities; so the fact that non-sampling errors cannot be easily calculated will not be relevant to this demonstration. The point to make is that such errors exist.)

Sample size	*Cost of investigation*	*Error (Percentage)*		
		Sampling	*Non-sampling*	*Total*
20	10	70	0.01	70.01
40	20	50	0.03	55.03
100	50	30	0.10	30.10
500	250	15	0.50	15.50
2500	1250	7	1.50	8.50
5000	2500	3	2.00	5.00
25000 (i.e. complete census)	12500	0	5.50	5.50

Attention should be drawn to the total error in the two cases corresponding to sample sizes 2500 and 5000 and comparison made with that for complete census. In the former case, results which are only a little more in error than the census results are obtained at one tenth of the cost. But in the second, even better results than in the census are seen, the cost being only one fifth.

8. *Exercises*

Examples:

(1) The use of random number tables for simple random selection. The difference between sampling with replacement and without replacement will be brought out; so also the need for suitable adjustments to reduce the rejection of random numbers selected from the table. At least the four cases typified by the following examples should be covered: (*a*) selecting a random sample of size 15, from a population of 1852 listed units, (*b*) selecting on a map an area 1/10 kilometre × 1/10 kilometre, at random, from an irregular area falling within a rectangle 4 kilometres × 3 kilometres, (*c*) selecting with a random start every seventh of the houses in a street, numbered 1 to 78, (*d*) selecting at random 5 houses from all the houses in 6 streets containing 52, 17, 32, 48, 29 and 14 houses respectively.

(2) Arrangement of given sets of characteristics into sections of data sheets, and giving necessary indications of the manner of filling in each item.

9. *Training material*

Lecture notes. Exercise sheets. Sampling schemes of selected investigations. Specimen data sheets. Demonstration material (section 6).

TOPIC 2.2

Details of designing

1. The designing of an investigation requires the working out of several technical and administrative details. These details will be based on decisions taken about the different methods and operations to be adopted in the investigation. Many of the decisions will be taken in accordance with an overall view and the details will be co-ordinated into an efficient system. The details would vary in nature from investigation to investigation. Broadly they may be considered under the following heads.

2. *Formulation of objectives*

A simple and clear statement of the objectives will be helpful to avoid the objectives being lost sight of in the complexity of details of designing and making decisions at variance with the objectives. Before the purpose of the investigation can be translated in terms of the statistics required, discussions with persons or agencies who initiated the investigation will be often necessary to clarify various issues. This will help to clear up vagueness and to arrive at decisions with regard to need, scope and coverage of the information, as also the sources, methods, cost and time required for obtaining the information. Sometimes it may be desirable to obtain additional information indirectly related to the objectives, at marginal cost. The objectives may also specify whether the investigation will be an isolated one or periodically repeated.

3. *Administrative preparation*

(1) *Budgeting:* In order to arrive at the total cost of the investigation and to discover what expenditures will be required according to a time schedule, it will be necessary to work out cost estimates for as detailed a breakdown as possible. Thus, for example, detailed cost estimates will be required, classified broadly under: (*a*) salaries, rent, equipment, material, travel and other heads, (*b*) the different stages of the investigation such as organization, designing, data collection, data processing, presentation and publication, (*c*) different periods in the time schedule. Cost estimates and budgets provisionally prepared may have to be revised in the light of experience and from the point of view of optimum allocation of resources.

(2) *Organization:* Necessary administrative arrangements for controlling, co-ordinating and supporting the various activities concerning the investigation may be set up either by fixing responsibility on one or more existing offices or by establishing a separate office. Designing in this context will be concerned with defining the functions of the administrative units and individuals, allocation of staff, equipment and material, and drawing up time schedules. The preparation of organization charts and an investigation calendar will facilitate administration.

(3) *Legal provision:* Consideration should be given as to whether adequate legal provision exists for the investigation. If not, it should be secured as an enactment, decree, executive order or the like, so as to achieve the ends in view without incurring hostility or non-cooperation. The responsibilities and bindings on the staff in respect of the legal provision should be made clear to all concerned.

(4) *Publicity:* For conducting investigations the success of which depends on co-operation from respondents or the general public, proper publicity must be given to the investigation, mentioning the purpose, the need for co-operation, the safeguards established to protect personal and public interests and the manner in which co-operation can be extended. Preparation of publicity material and arrangements for communicating it through pamphlets, newspapers or the radio will be an important aspect of designing. Publicity material and the arrangements for using it will have to be prepared in the light of prevailing conditions. Thus urban and rural areas may require separate treatment.

(5) *Personnel:* The personnel requirements for various items of work involved in the investigation will have to be laid down in terms of the number of persons needed and their qualifications for eligibility. Recruitment of staff may require issuing invitations for applications and conducting selection tests. Training of staff for different types of duty, particularly for field work, processing and the like will need special organization.

(6) *Equipment and facilities:* The different items of equipment, material and facilities required for the investigation will have to be listed and procured. Plans for their location, distribution and control will have to be prepared. The equipment should include instruments and machines for operations such as data collection and processing and transport vehicles.

4. *Technical preparation*

(1) *Concepts and definitions:* The designing activities under this head should include first the preparation of a list of concepts concerning the characteristics and entities to be covered by the investigation and then the provision of clear definitions of: (*a*) the population and its characteristics; (*b*) the individuals and their characteristics; and, if sampling is adopted, (*c*) the sampling units, which may or may not be different from the individual.

The concepts and definitions may be laid down in conformity with national or international standards or may be specially prepared to suit the particular investigation. Special attention must be given to ensure that the characteristics selected are valid and most appropriate for measuring the concepts, that the coverage with respect to the individuals is complete and that time references and units of measurement are, where necessary, clearly specified in the concepts and definitions.

In some cases, as in bulk material, a self-determined or integral individual or sampling unit may not be available. The appropriate 'size' of the individual or unit will then have to be separately determined beforehand on the basis of past experience or pilot data.

It is necessary to verify that all the data covered by the lists prepared are relevant to the purpose of the investigation and that no essential data are omitted.

(2) *Data collection:* Depending on the sources from which the information will be collected, the nature of information and facilities available, the method and details of data collection to be adopted will be laid down. The characteristics to be observed will then be considered individually or in sets, with regard to the methods of observation to be used. Adequate information and material will thus be available for preparing the data sheet and instructions for data collection. When the data sheet and instructions have been prepared, a 'try-out' of the data sheet may be made, especially when field work is involved, using selected staff as well

as those staff who are later to collect the data for the investigation. The data sheet and instructions will be finalized after the try-out.

(3) *Selection of individuals:* A complete and up-to-date frame of units will be necessary in many investigations to enable the selection and location of individuals for observation. Lists of individuals, households, villages or towns, maps of rural areas, town plans and aerial photographs can serve as frames. Frames may be constructed, if they are not already available, from records or by carrying out field work. If sampling is adopted, a selection of sampling units is made from the frame in the manner required by the type of sampling adopted, and a list of sample units prepared. In some situations, it may not be possible to prepare a frame. In that case special selection procedures will have to be prescribed.

(4) *Tabulation:* The determination of the layout and content of the final tables to be obtained as a result of processing the data is an important item to be carried out at the designing stage itself so as to ensure that the data collected will definitely lead to results consistent with the objecives of the investigation.

(5) *Control of quality:* To ensure the quality of results, checks may be introduced at various stages of the operations. *For controlling the quality of data collection,* methods such as inter-penetrating samples and post-enumeration checks may be used. If the mail enquiry method is adopted, the procedure for 'call back' should be formulated in order to get information from non-response cases. In processing work, checks of different types for detecting numerical errors may be introduced.

(6) *Training:* Preparation of training manuals and other material for imparting training on principles and techniques involved in the various operations will be an important item in designing, especially when new staff is recruited or when a new investigation is undertaken.

(7) *Instructions:* Apart from any time-schedules prepared for the different stages of the investigation, it will be necessary to make a full review of the different operations and procedural details under each operation. Schemes will have to be prepared for each operation and the staff instructed as to how to carry out each item of work under the operation. The instructions should be complete as regards the background knowledge required, techniques to be adopted, methods of using instruments and other equipment, precautions to be taken and administrative rules to be followed. Separate instruction manuals may be prepared for different operations such as field work, compilation from records, scrutiny and editing of data sheets, computation, tabulation, charting and publication. Instruction manuals may also contain material for reference, such as concepts and definitions, classifications, code lists, specimen administrative forms, table of random numbers and conversion tables.

Training on Topic 2.2

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Consideration of the results of designing activities with reference to a selected investigation.

6. *Demonstration*

Example:

Details of designing. While, in the demonstration under Topic 1.5, attention was confined to operations under a single investigation, in this one, all aspects of designing should be covered and therefore material and information relating to more than one investigation may be necessary. Both administrative and technical details of designing should be dealt with by presenting selected results of designing. The material presented and distributed to trainees should include, specimens of different types of: budgets, survey calendars, publicity material, training schemes, lists of equipment used, schedules, questionnaires, instruction manuals, scrutiny plans, processing plans, computation plans and presentation plans. Information about different types of legal provisions, publicity arrangements, agencies responsible for the conduct of the investigations, administrative organization, sampling schemes, data collection methods, processing methods and publication arrangements, should also be provided.

7. *Exercise*

Example: Preparation of scheme of tables that can result from information collected on copies of a given data sheet. The data sheet should be brief. The trainees need indicate only the classification and contents scheme of the table and not prepare the full blank forms for each table proposed.

8. *Training material*

Lecture notes. Exercise sheet. Copies of the design of a selected investigation. Demonstration material (section 6).

Course 3

DATA COLLECTION BY FIELD WORK

This course is for trainees at the primary level and the intermediate level. Its purpose is firstly to create an awareness about the qualities that a field worker should possess, then to provide some background knowledge on some items that may be found useful in field work relating to agricultural, industrial and other socio-economic investigations, and finally taking the trainees through the different items of field work.

Training will be given through lectures, discussions, demonstrations, exercises and visits. One of the useful modes of training under this course will be to make the trainees watch details of different items of field work being demonstrated by instructors and by professional field staff.

Course summary:

Topic	*Number of hours proposed for instruction*				
	Lecture/ Discussion	*Demons-tration*	*Exercise*	*Visit*	*Total*
3.1 Field workers' requirements	6	2	4	–	12
3.2 Field procedures	10	6	8	4	28
TOTAL:	16	8	12	4	40

TOPIC 3.1

Field workers' requirements

1. Field work may be done either as the chief means for collection of primary data or to supplement other methods of data collection, as for instance to obtain information from non-response cases in a mail enquiry or to fill up gaps in data compiled from administrative records.

In data collection by field work, trained staff visit the spots where the individuals are located, and make the observation. Any type of data sheet and any method of observation may have to be used. The individuals observed may be persons, households, industrial establishments, farms, points or lines or areas on land, markets, traffic junctions, industrial products, animals, fish, trees, fruits, portions of solid or liquid or gas, periods of time or other entities. The observation may be aided by instruments for measuring, counting or recording or by other equipment. In some cases the instruments used may be automatic or self-registering.

The reliability of statistics obtained from data collected by field work will depend on the accuracy of execution of field work. Indifferently collected data will be subject to biases and other errors and the resulting statistics may be grossly misleading.

2. *General requirements*

The field worker will almost invariably have to carry out his activities away from his office or headquarters. He will be provided with detailed instructions regarding concepts, definitions and procedures concerning the data collection besides the necessary instruments and other equipment. He should be able to fully understand and act according to the instructions and to make efficient use of the given equipment. He should take the opportunity to clarify all doubts and questions at the time of training and pilot work. In the collection of accurate data within the prescribed time schedule, his other equipment is his own knowledge and personal qualities such as clearness of understanding, sense of accuracy, freedom from biases, neatness in handwriting and number work, ability to take precautions, tact, resourcefulness, patience, honesty. Also he should have good health and be alert and active.

3. *Knowledge requirements*

There are several items of knowledge that would help in the execution of field work and contribute towards the quality of data collected. They differ according to the field of operation. So also the depth of knowledge required depend on whether the work is one of primary data collection or one of organization and supervision of field work. Some of the special items of knowledge useful for a field worker who may be called upon to perform duties in different types of statistical investigations are given below.

(1) Important aspects of the geography of the area concerned—administrative divisions, their boundaries; location of towns and their importance; general aspects of climate; location of mountains and forests; courses of rivers; railway routes; location of railway stations, seaports and airports; natural resources, crops, industries, industrial organizations, industrial projects.

(2) Travel and postal information—rates and rules prevalent in the area concerned.

(3) Administrative structure, including village administration within the country, executive officers and their functions.

(4) Weights and measures, systems and conversion from one system to another.

(5) Concepts and definitions generally adopted in statistics, such as those relating to: household, family, household income, individual occupation, individual income, earned income, gainful activity, labour force; public sector, private sector, joint stock company, co-operative society, fixed capital, working capital, input, output, hired labour, exchange labour, household labour, agricultural labour, repair and maintenance, wages and salaries, capital formation; small scale industries, large scale industries; land owned, land leased in, land leased out, land possessed, fallow land, gross area under crops, net area under crops, area sown, area harvested, area irrigated, mixed crops; producers' price, consumers' price, wholesale price, retail price, harvest price.

(6) Sources of lists of villages or rural areas, lists of households, lists of holdings, lists of blocks in towns, village or rural maps and town maps.

(7) Elements of geometry leading to understanding of scales of maps, measurement of distances and areas on maps, compass directions, elevations and gradients of land. Also familiarity with standard symbols used in maps, such as those for railway stations, roads, shrines, schools.

(8) Use and care of instruments such as weighing balances, stop-watches, counters, planimeters.

Training on Topic 3.1

4. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 3 above.

(2) Consideration of specimen lists of duties of field workers, relating to selected statistical investigations.

(3) Instruction on items, selected from section 3 above, according to the needs of trainees attending the course.

5. *Demonstration*

Example:

Interpretation of instructions for field work. The purpose of this demonstration is to give a general introduction to the problems in field work. For this demonstration, a typical but brief set of instructions for field work together with the corresponding data sheet will be chosen. Copies of the instructions and data sheet will be made available to each trainee in advance of the demonstration, for study.

The background of the investigation will be briefly described. The instructions will be discussed, and all aspects of work will be explained indicating procedures involved in arriving at each entry in the data sheet, to ensure that the instructions have been so well grasped by the trainees that they will be able to do the field work themselves.

6. *Exercises*

Examples:

(1) (*a*) Copying maps, on reduced scale, involving use of symbols, and measuring the area of a bounded region of the map (using planimeter or transparent graph paper).

(*b*) Preparation of lists for guidance in field work (persons, households or establishments).

(*c*) Preparation of sketch maps for guidance in field work.

(2) (*a*) Fixing, at random and systematically, points, lines and areas on maps.

(*b*) On a picture map of partially cultivated land, in which shading is preferably used to indicate crop, fixing points, lines and areas and using them to estimate the proportions of area under crop by (*a*) counting number of points, (*b*) measuring intercepts of lines and (*c*) measuring areas, falling in the portion under crop (shaded).

7. *Training material*

Lecture notes. Exercise Sheets. Specimen lists of duties of field workers. Copies of selected instructions for field work and copies of corresponding data sheets required for demonstration (section 5). Maps, picture maps and other material required, depending on exercises set (section 6).

TOPIC 3.2

Field procedures

1. Mainly, the items of work that will have to be performed by field staff are the following:

(1) Locating and identifying individuals, (2) Listing units, (3) Selecting individuals, (4) Making the observation, (5) Checking recorded information, (6) Returning completed data sheets.

Each of these items is briefly described below, giving some general points for guidance of field staff in each case.

For any particular investigation, the instructions will lay down the different concepts, definitions and procedures involved, together with detailed guidance regarding each aspect of field work. A prerequisite to field work is a careful and detailed study of the instructions and obtaining necessary clarification before leaving for the field.

2. *Locating and identifying individuals*

This is an essential step before making observation. The individuals may be located in space, as for example: houses, factories, villages, or located in time as for example: time periods concerning vehicular traffic across a point or amount of fish unloaded at a landing centre. Locating and identifying each individual may be done with the help of a set of definitions, a list, or a map which will give the necessary indications. In some cases, even after locating a spot with the help of the list or map, a further selection of the individual will have to be made by the field worker in a manner laid down in the instructions. Locating and identifying units or individuals will also be necessary if the field worker is to collect information first for preparing a list or a map. For that purpose a systematic procedure using all available information will have to be adopted. In all cases, the field worker should be able to recognize the individual or unit unambiguously and physically identify it in the field of operation. For example, he should have a clear idea with regard to the distinction between a house and a household, between a radio owner and a radio listener, between working hours and other hours and so on. He should be able to identify on the ground the boundary of an area given on the map with the help of landmarks, and similarly he should be able to trace on the map the physical boundary of a given area seen on the ground such as a cluster of fields or a block of houses.

Clear definitions about time intervals or other units for observation, lists, sketch maps showing boundaries of fields, location of dwellings and the like, notes regarding distances, directions and other indications, and plans regarding mode and time of travel, prepared in advance, would facilitate the correct location and identification of units and individuals for observation. In rural areas, the local knowledge of inhabitants may be of great help in identifying boundaries of villages and fields.

3. *Listing units*

When a list of units or individuals is not already available and cannot be made from sources like records and registers, it will have to be prepared as a result of field work. Such a list may be required for providing a frame for selecting sample units or as an aid to the field worker in locating and identifying units at the time of observation.

In preparing such lists, care should be taken that no gaps are left in them anywhere. When an existing list has to be brought up to date, a systematic verification of each unit is necessary. The extension of town boundaries or construction of new houses often renders house lists out of date. In preparing a list of manufacturing establishments, for instance, care should be taken to include new establishments and omit inoperative ones. Eligibility of doubtful units should be verified, as for example, whether a structure is a house or not. Items in the list should preferably be ordered according to the route taken while preparing the list, with indications of the direction of the route.

4. *Selecting individuals*

In many investigations, the selection of units will have already been made at the designing stage, but sometimes the field worker may have to select the individual for observation after reaching the spot of observation; for example, he may have to choose one of two halves of a patch of crop to be harvested, or every fifth dwelling unit in a street, with a random start. Where units are to be selected at random, the selection will have to be made by tossing coins or dice or using random number tables. Random selection is not haphazard selection, but a well defined objective procedure.

If for any reason it is not possible to observe a selected unit, substitution should not be made unless a procedure for substitution is given in the instruction or without obtaining specific instructions from supervisory staff. Skill in field work lies in obtaining information from all selected units, except when it is physically impossible to do so, as when a selected household has left station permanently.

5. *Making observation*

Observation may be made by measurement or counting or eye-estimation and personal judgement or by taking answers based on interviews. Where observation involves measurement, the units of measurement or the accuracy laid down in the instructions should be maintained. If local units of measurement are employed, they should be converted to the standard units at the appropriate stage. Use of local equipment for measurement such as balances and weights should normally be avoided and the field worker should be carrying a kit containing instruments and other equipment. Measurement and counting are most desirable forms of observation, from the accuracy point of view, but occasionally observation may have to be made by eye estimation or personal judgement, for example, estimation of the number of sheep in a flock or the proportion of crop area in a field.

When the observation involves interview with a respondent, the worker should approach with due respect for the respondent, and not in exercise of any official powers that he may have. He may produce his credentials and explain the objective of his visit in a language which the respondent should appreciate, and indicate, if necessary, that the information given will remain confidential. Knowledge of local conditions and customs is a great advantage in creating the right atmosphere. The time chosen for the interview is equally important. Interviewing has often to be done in the evening hours. The questions asked should not be of a leading character. They should be asked strictly in the order in which they appear in the data sheet unless separate instructions have been given. In making questions and in obtaining answers, attention to wording is important, for instance, in a question or answer regarding the work done by a person, attention should be paid to the reference period mentioned (day, week, or month).

In entering the results of observation on the data sheet, special attention should be given to instructions regarding the use of standard symbols for answering yes-no type of questions, abbreviations such as kg., cm., hr., when to write with pencil and when with ink, when to fill the data sheet in the field itself, and when to fill it later and so on, so as to ensure uniform entries by different field workers.

6. *Checking recorded information*

Errors can occur at the time of observation in several ways. At the time of recording each entry and after each section of the data sheet has been filled, a double check will always be useful. The field worker may sometimes have to make more than one visit to the place of observation either for making repeated observations or for a purpose like thanking and bidding farewell to respondents. He should take such opportunities to confirm the accuracy of or correct the data previously collected. In any case, the checking has to be done while the field staff is still in the field. In some data sheets, there will be provision for making cross-checks. This will enable the mutual consistency of the recorded information to be verified; for example, information on size of holdings, land utilization, number of ploughs and cattle possessed should be mutually consistent. Sometimes data sheets are filled in after some auxiliary notes have been made. The entries in such notes and any calculations made in them or in the data sheets should also be checked.

7. *Returning completed data sheets*

The data sheets duly filled and checked for completeness, dated and signed, should be despatched to the controlling office with the minimum of delay, together with information sheets such as diaries of work and travel. Discrepancies noted at the controlling office will be referred back to field staff for verification.

Training on Topic 3.2

8. *Lectures and discussions*

(1) Presentation of the contents of sections 1 to 7 above.

(2) Consideration in detail of the different items of field work (section 1) in a particular statistical investigation suitably selected.

9. *Demonstrations*

Examples:

(1) *Interviewing respondents.* The purpose of this demonstration is to acquaint the trainee with the art of interviewing respondents. The first part of the demonstration will consist of enabling the trainee to listen to selected interviews being made. This may be done by playing recorded tapes corresponding to a given questionnaire. Four different types of interviews may be demonstrated (1) with the head of a household (male), (2) with the housewife, (3) with a farmer, and (4) at an industrial or commercial establishment.

Whether recorded tapes are available and played or not, the trainees should be given the experience of actually doing the interviews themselves. The directed role-playing method should be adopted for the purpose.

The trainees will take turns in playing the part of the investigator and of the respondent. As this is a time-consuming process, extra time may have to be allowed for this demonstration. Further practice in interviewing should however be left to be gained from doing the exercise (Example 4 below).

(2) *Locating and identifying units.* This demonstration will be held in the field and precede the exercise on the subject (Example 1 below). The trainees will be taken out for field work, if necessary, in batches under different instructors. They will be shown how units are to be located and identified on the basis of definitions and a list or a map. Necessary lists and original or sketch maps and plans for reaching the units should be prepared beforehand. Location and identification of units like dwelling units, establishments or farms on the basis of lists and units like fields, on the basis of maps should be covered. Selection of the individual for observation, by a random or other process, after locating a unit should

also be included. Such individuals may be persons, points, lines or areas on land or a section of bulk material. This demonstration should also include the identification on maps, units shown on the ground, identification of different crops (including mixed crops) and trees, and the preparation of sketch maps carrying information for later work.

(3) *Making observation in the field.* This demonstration will be held in the field and precede the corresponding exercise (Example 2 below). The method of field observation by interview is covered in Demonstration 1 above. This demonstration will include the other types of observation in the field. In practice, field work will often involve measurements using instruments, such as measuring the length of corn ears, weight of grain, weight of fish or other food stuff, or determination of crop yield in a geometrically defined 'cut' or patch of crop. Reasonably adequate demonstrations of such measurements can be given in the classroom or during visits. This demonstration should give preference to approximations to measurements such as eye-estimation of weights and determination of lengths by counting steps. The items that should be illustrated during the field demonstration include: *i*) estimation of total area under crop, *ii*) proportion of area under crop, *iii*) yield rate of crop, *iv*) distance between two land marks, *v*) distance along a road, *vi*) weights of objects, *vii*) volumes of liquids.

10. *Exercises*

All the exercises under this topic will be outdoor exercises, carried out in the field and may follow corresponding demonstrations.

Examples:

(1) (*a*) Identification on ground, units such as plots of land and dwelling units, shown on map.

(*b*) Identification on map, units shown on the ground.

(*c*) Identification of crops and trees shown.

(*d*) Preparation of sketch maps containing information collected during field work.

(2) (*a*) Locating specified points, lines and areas on ground.

(*b*) Estimation of: *i*) total area under crop in a farm or a field, *ii*) proportion of area under crop in a farm or field, *iii*) yield rate of crop in a field and *iv*) yield rate of crop in a demarcated patch of crop.

(*c*) Estimation of: *i*) distances between land marks, *ii*) distances between points along a road, *iii*) weights of objects, *iv*) volumes of liquids.

(*d*) Estimation of a characteristic relating to bulk material (examples: *i*) impurity in coal or sand, *ii*) yield of fruit on a tree by selecting at random the unit for observation, after successive dichotomies.

(When eye estimations are to be made, a number of trainees should be given the same objects for estimation and provision made for comparison of estimations by different trainees, whenever possible with actual measurements).

(3) Listing of all dwelling units in a portion of a village or a street, with full identification details, and selecting a stated fraction of houses, *i*) systematically with a random start, and *ii*) at random.

(4) Filling in by field work at least 3 copies of a given data sheet. The data sheet should be one requiring the interviewing of respondents.

11. *Visits*

(1) A field organization engaged in socio-economic enquiries, involving visits to households and interviewing respondents, or an organization engaged in collecting data from establishments by field work.

(2) A field organization engaged in collecting agricultural data. Preference may be given to a place where crop-cutting or a similar measurement type of observation is conducted.

12. *Training material*

Lecture notes. Exercise sheets. Descriptive account of field work in a particular investigation (section 8 above). Demonstration material (section 9 above). Exercise material (section 10 above).

Course 4

DATA COLLECTION FROM RECORDS

This course is intended for both the levels, primary and intermediate. The purpose of the course is to familiarize the trainees with problems of routine compilation of statistics and to provide knowledge about the details, principles and methods involved in the production of statistics of good quality from information contained in administrative documents or routine statistical returns or published or other statistics already available.

Training will be given through lectures, discussions, demonstrations and exercises. To the extent possible, the material for compilation and compilation schemes referred to or used for illustration during training should relate to the statistical offices where the trainees will work after training.

Course summary:

Topic	*Number of hours proposed for instruction*			
	Lecture/ Discussion	*Demonstration*	*Exercise*	*Total*
4.1 Material for compilation	6	2	2	10
4.2 Compilation schemes	6	2	4	12
TOTAL:	12	4	6	22

TOPIC 4.1

Material for compilation

1. Data collected from records may consist of information available in publications, data gathered from administrative records and data from routine statistical returns. Such data form the basis of most statistics compiled as a routine. The information will often have to be transferred to suitably prepared data sheets before the necessary statistics can be produced. This type of data collection presents some problems different from those of data collection by a specially designed investigation. The difference mainly lies in the fact that the scope and coverage of the information are almost fully determined by the records from which information is extracted. There will, however, be no appreciable differences as regards methods of processing the data and presentation of the results. Since statistics compiled as a routine are used for purposes which would in their absence have been served by statistics produced by special investigations, special treatment of the former will be necessary both in respect of their compilation and their utilization.

2. *Importance*

Statistics of a wide range and variety can be compiled from data taken from returns, registers and other records prepared in the course of administration, and from publications. A major part of the statistics produced by governments, private and public concerns, business associations and international organizations is based on data taken from internal or external, statistical or non-statistical records. Such statistics are mostly compiled as a routine and are intended to serve various current and future purposes.

An important feature of collecting data from records is that the cost of collection can be considerably lower than in the case of special investigation as the data are obtained as by-products of administration or as secondary statistics.

Some of the records, besides providing data to produce statistics, furnish the frames necessary for conducting special statistical investigations.

Governments, public bodies and many other organizations keep various types of records containing valuable statistical information, some of which may not have been utilized to produce statistics. Thus there will often be the possibility of making increased use of records for producing much needed statistics at low cost.

There is also the possibility of producing more and better statistics by making suitable modifications in existing systems of records.

3. *Types of material*

Two important types of administrative records from which data are extracted for compiling statistics are: (*i*) documents maintained, received or issued on a regulatory basis, such as registrations, applications, permits and licenses and (*ii*) records relating to internal activities of institutions such as progress reports and expenditure statements. Examples of the former type are: birth and death registers, business licenses, labour returns, tax returns, customs declarations. Examples of the second type are: public revenue and expenditure statements, employment exchange records, business records of firms, operating records of social insurances, public corporations and government enterprises.

Another important source of material for compilation is statistical returns submitted by different organizations to larger bodies. The larger bodies assemble the statistics contained in such returns and make them available in a consolidated form after standardizing them as far as possible. Examples of statistical returns of this type are those relating to production, sales, employment, etc., sent by individual firms to business associations or those sent by provincial government offices to regional or central offices or by national governments to international agencies.

Yet another source of material for compiling statistics comprises publications by individual authors, private or public institutions, governments and international bodies.

4. *Limitations and improvement of quality*

Since most of the records yielding material for compilation, have primarily to serve purposes other than producing statistics, there are likely to be limitations in the material, from the statistical point of view. Thus, for instance, concepts and classifications followed in the records may be different from those most desirable for statistical purposes. Figures of the same type from one source may differ in definitions, classifications, units of measurement, accuracy and so on, from those from another source. The coverage of the data may be indefinite or incomplete. The information available may be imperfect on account of reluctance, under-statements or exaggerations on the part of those responsible for the information in the records. There may be delays in reporting or in making available the records for compilation.

There are various ways in which records can be improved so as to yield comparable and reliable statistical information, without seriously affecting their administrative uses. This however, requires a careful study of, among other things, the administrative functions of the records, the regulatory provisions, if any, the sources and channels of the records, concepts, definitions and classifications used, the design of the forms used, the timing and methods of completing the records and the degree of co-ordination between related types of records.

To ensure that the coverage is complete an examination to discover whether all the units and individuals are clearly defined and included in the records will be useful. Concepts and definitions in use may not be standard or may be out of date and may require revision. If a revision cannot be effected for some reason, ways to convert one definition to another or indications of how two concepts differ should be provided. Combining characteristics or individuals diversely defined, into broader groups and reclassification of material may increase the utility of the resulting statistics. Efforts to keep records up to date, providing clear time-reference, avoiding overlapping periods and gaps and prompt release of records can greatly improve the quality of compilation material. While re-designing the format of records it may be possible to include additional characteristics without creating inconvenience to those responsible for completing the records. Sometimes, useful information is not available because it is labelled 'confidential'. Periodic review of the need for continuing to clasisfy it as confidential may be made in view of the uses to which such data can be put. In addition, the creation of a consciousness on the part of those who maintain the records that records are the raw material for important statistics and acquainting them with the basic statistical requirements of the records, such as unifomity, continuity, completeness, accuracy and timeliness of information, can contribute to the quality of compilation material. For this, training of the staff who keep the records and framing instructions for completing the records will be found useful.

Training on Topic 4.1

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Consideration of features of different types of material for compilation, with a view to understanding their background and to determining the statistical information that can be extracted from them, their adequacy to meet specified purposes and the improvements possible.

6. *Demonstration*

Example:

Differences between data from records and data by investigation.

An idea of the nature of data collected by special investigations will already have been given through the demonstrations under Topics 1.5 and 2.2 For the purpose of this demonstration, specimen administrative records, e.g. registration forms, licences, records of employment exchanges, public corporations, published statistics of prices or employment and returns such as those relating to labour and tax should be made available. The limitation of the records from the point of view of extracting data for producing specified types of statistics should be pointed out. The corresponding data sheets used to transfer the information from the record, return or publication should also be presented and discussed. The need for appreciating the limitations of the sources should be indicated. The differences in scope and methods associated with statistics derived from records and existing primary statistics, as compared with statistics based on special investigations should be brought out. The material used for the demonstration should as far as possible be related to offices with which the trainees are familiar.

7. *Exercise*

Example:

On the basis of specimen forms of: (*i*) an administrative record and (*ii*) a statistical return,

(*a*) preparation of a list of statistical characteristics that can be compiled out of information collected in them.

(*b*) preparation of a note as regards improvements that can be made in them and listing the benefits arising from such improvements.

8. *Training material*

Lecture notes. Exercise sheet. Specimen copies of material for compilation, such as administrative records, returns and published statistics. Lists of concepts and definitions relating to the material for compilation. Copies of instructions regarding completing and channelling records and returns. Copies of data sheets used for transferring information from the original material.

TOPIC 4.2

Compilation schemes

1. The main problems in preparing a compilation scheme are to visualize the end results that can be obtained from a given set of records and to make the best use of the available information to obtain the desired results in as short a time as possible. The administrative and organizational details of compilation work would vary according to the size of the work and the type of agency responsible for it. In some cases the same administrative agency which is responsible for producing the records may itself compile the statistics for its own publication or for the submission of statistical returns to another agency. In other cases, copies of the basic records will be passed on for compilation to an office specializing only in statistical work. Designing compilation requires attention almost similar to that for designing a special investigation and the administrative and technical aspects of data collection. processing, presentation and publication will be covered by the design (Topics 2.1 and 2.2).

2. *Study of material to fix end results*

The first step in designing a compilation scheme is to make a study of the available records with a view to determining the statistical results that would be possible out of the data. This will have to be done in the light of any statistics that have been produced in the past from the same type of records and of the possible uses of the statistics that will be produced. Possible improvements in the quality of the resulting statistics will also be kept in view.

For this purpose consultation with the producers of the records will be necessary to obtain background knowledge about the records, such as purposes for which the records are maintained, the parties responsible for providing information contained in the records, the administrative channels and procedures the records have gone through, and the meanings of terms used in the record. So also consultation will be necessary, with the possible users of the statistics that will be produced, to avoid omission of important aspects and inclusion of unnecessary details.

The information contained in the records will have to be visually analyzed in terms of suitable population characteristics and individual characteristics from which the former can be built up. This will reveal gaps and other limitations in the information contained in the records. It may be possible sometimes to minimize gaps and other imperfections, by considering a slightly different set of characteristics and individuals. Careful consideration will have to be given to the question as to whether the resulting statistics will really give rise to the desired population characteristics or whether can at best only yield sample characteristics. In some cases, gaps and limitations may be overcome, by introducing a suitable sampling design.

A study of the above type should result in a decision in respect of the nature of information to be selected from the records and the manner of selecting material for transfer to data sheets.

3. *Data sheet*

Unless punched card methods are used, the information in the records may have to be transferred to data sheets. In most cases, simple schedules or forms will suffice. However, the usual principles of preparing data sheets and instructions for filling them will have to be followed. The process of transferring information will in most cases consist of posting information selected from records, in appropriate classification cells on the data sheet. One of the considerations, therefore, in the design of the data sheet is that the classification characteristics chosen should be positioned on the data sheet, in such a way as to facilitate both transfer of data as well as subsequent processing work. Thus for instance, sometimes data from more than one type of record will have to be transferred to the same data sheet. In that case, the classification characteristics will be so grouped as to suit the sequence of entries from the different records. The position of a classification characteristic in a hierarchy is also important from the point of view of obtaining totals of entries made. It will be seen that a wrong position of a classification characteristic, would necessitate selective addition instead of continuous and direct addition for obtaining desired totals, when the data are processed manually. Provision will also have to be made on the data sheet for working out any adjustments in magnitudes, that may be necessary. For example, in customs declarations goods may be quoted at prices including delivery charges for which adjustments will have to be made.

When punched card methods are adopted, the transfer of data from one or more types of record will be made directly to cards according to suitable card designs and the question of transfer to data sheets does not arise.

Before making entries on data sheets or punching information on cards, the records will be scrutinized for internal consistency, accuracy and completeness and rectification, replacement or rejection of information as necessary will be made.

A common function of statistical offices is to compile information from official returns sent by subordinate offices or forms filled by private individuals or organizations. The designing of the format and content of the return or form should receive special attention. Very often the return or form could itself serve directly as the data sheet.

4. *Processing, presentation and publication*

In spite of scrutiny of the records, errors and inconstencies not detected earlier may be brought up on the data sheets. There may also be errors made during the transferring operation. A scrutiny of the entries on the data sheet will therefore be necessary. After such scrutiny, the data may have to be summarized on a different set of working sheets under broader classifications. The summary will be checked for accuracy. Such summary tables will be prepared also when statistical results are extracted from publications. The next step may be the aggregation of data to obtain different subtotals and totals. The results of processing would lead to detailed tables. Tables so compiled on each occasion will generally give rise to magnitudes constituting terms in a continuous series. Further summarization and computation may be necessary to fit the results into an existing series. Some of the results may be charted. Where necessary textual notes will be added. The results of the compilation will then be released for printing and publication.

At the designing stage, therefore, scrutiny programmes, forms for summaries, table lay-outs, charting procedures, publication programmes, flow-charts, budgets, organizational details and overall time-programmes will be prepared (Topics 2.1 and 2.2).

Training on Topic 4.2

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Consideration of details of preparing a compilation scheme (covering the stages from the basic record to publication of results) with reference to a selected set of records or returns.

6. *Demonstration*

Example:

Compilation schemes. Material relating to three selected schemes of compilation will be presented and described: (*i*) compilation from records, (*ii*) compilation from statistical returns and (*iii*) compilation from published statistics. In each case the material will include, the source of data, the data sheets used, instruction manuals, and plans and programmes for scrutiny, processing, presentation and publication. The differences in operational details as compared to statistics collected by special investigations will be brought out.

7. *Exercises*

Examples:

(1) Preparation of form (data sheet) for transferring information from specimens given of:

(*a*) non-statistical administrative record,

(*b*) statistical return of an administrative character,

(*c*) published statistics.

(2) Scrutiny and transfer of information to given data sheets from the three types of sources. Preparation of notes on limitation, if any, of data.

8. *Training material*

Lecture notes. Exercise sheets. Specimen compilation scheme of a selected compilation project. Demonstration material (section 6 above).

Course 5

DATA PROCESSING

A major part of this course will be gone through by trainees both at the primary and intermediate levels. Items intended for those at the intermediate level are some of the more advanced computational and analytic techniques appearing in Topic 5.2, and are indicated separately.

The objective of the course is firstly to provide acquaintance with the principles and methods of large scale data processing and to provide practice on computational techniques. While some basic knowledge of statistical methods will have been imparted through Topics 1.6 and 1.7, any statistical method introduced during training on Topic 5.2 is intended to be mainly through computation.

Training will be given through lectures, discussions, demonstrations, exercises and visits.

Course summary:

Topic	*Number of hours proposed for instruction*				
	Lecture/ Discussion	*Demonstration*	*Exercise*	*Visit*	*Total*
5.1 Processing operations and principles	8	4	4	4	20
5.2 Computation techniques	12 (6)	2	26 (14)	–	40 (20)
TOTAL:	20 (6)	6	30 (14)	4	60 (20)

(The figures in brackets show hours to be spent by intermediate personnel only and form part of the figures below which they appear)

TOPIC 5.1

Processing operations and principles

1. Data processing consists of the various operations required to produce statistical results from information contained in data sheets. No new information can be created by processing. Data processing operations may thus include: scrutiny of data sheets, coding of information, transfer of information to working sheets or punched cards, classified arrangement of data, counting for class frequencies and arithmetical operations on the data to obtain summary results. A number of principles, techniques and equipment are involved in data processing operations. While data processing methods vary according to the quantity of data to be handled, the depth of statistical analysis and the processing equipment available, they do not depend so much on the method of data collection.

2. *Planning*

A number of decisions will have to be made and details worked out before processing operations can start. Procedures and arrangements will have to be determined for receiving, storing and handling data sheets. Control should be established on the number or quantity and location of data-sheets during different stages of processing. Special arrangements will have to be made for protecting data sheets or punched cards from dampness and other causes of damage; and for storing them so that they are easily accessible and released for processing work.

When the total volume of data sheets, the results to be arrived at, and methods of processing (manual or mechanical) are decided upon, details of the different operations involved, such as, time estimates, the sequence of operations, the allocation and flow of work, personnel and equipment requirements and cost involved are worked out.

So also, code lists, card designs, scrutiny programmes, flow-charts, lay-outs of working sheets, summary sheets, tabular forms, as necessary, are prepared. Besides, schemes and procedures for control of accuracy of results, for rendering processes fool-proof, for ensuring uniformity in operations, for achieving overall economy, and for conformance to time-programmes are laid down.

3. *Processing operations*

(1) *Scrutiny and editing of data sheets:* Data sheets received and properly stored will have to be first "cross-examined" and edited before the information contained in them is processed. The purpose is to detect gaps, inconsistencies and other imperfections and to rectify them so as to obtain reasonable quality levels, through background knowledge, internal evidence, or by reference to a data collection agency. Some of the common defects

arise from non-response, defective response, incorrect entries, illegible entries and omissions. Conformance to instructions for data collection is an important aspect to be verified. For any particular data sheet, it will be possible to prepare in advance, a list of likely errors giving methods of rectifying each error, and to scrutinize and edit data sheets systematically with the help of the list.

Checking for reliability of information should be extended through subsequent processing operations as well, apart from checks on errors in the operations themselves. Consistency between subtotals and totals, and between associated characteristics, and abnormalities such as outlying values and undue concentration of values, may be looked for during different operations.

(2) *Coding:* some of the entries in data sheets may have been already coded for convenience in making entries as also to facilitate processing work. Further coding may still be necessary. The purpose of coding at the processing stage is to add to the convenience in handling information, particularly at the stage of transferring information from data sheets to working sheets or punched-cards. Thorough acquaintance with code lists, the classification systems therein, and the codes themselves and practice in coding will be necessary to attain a reasonable amount of speed and accuracy in coding.

(3) *Transferring of information:* In order to facilitate the operations of sorting, counting and computation, in most cases the information on data sheets will have to be transferred to working sheets or special types of cards. Transferring may also have to be done to obtain duplicate sets necessary to safeguard against damage by use or for simultaneous distribution to different sets of workers or machines.

(4) *Sorting:* The sorting and arrangement into groups or classes, of data sheets or of cards on which information has been transferred, is an essential step towards performing group calculations and for obtaining classified information. In many problems, sorting and re-sorting the material according to different classification criteria, sorting in a hierarchy according to a set of criteria, matching and merging different sets of sorted material, or selection of particular groups from sorted material will be necessary.

(5) *Arithmetical operations:* Summarizing information, in a classified or other manner, involves the operations of counting, addition, subtraction, multiplication and division. Counts giving the number of cases in different groups maintained systematically will serve as an efficient means of exercising control over the accuracy of different processing operations, even if they do not have to be finally presented.

4. *Manual methods*

When the amount of processing work is relatively small, as when there are only a few data sheets to be handled, manual processing methods are economical. Operations of arrangement such as sorting data sheets into classes, merging, matching or filing data sheets in some sequence, or selection of a specified class of data sheets can be done manually when the number is limited. The counting of data sheets belonging to different classes, or copying values from data sheets to working sheets as well as arithmetical operations on the values copied out can also be done manually.

There are however several aids to manual data processing. For sorting purposes, *marginal cards* may be used. The positions of pre-punched holes along the edge of the card represent the different alternate forms of characteristics to be observed. One card is ordinarily used for each individual observed. An observation of a characteristic is registered by punching a slot opening outwards from the corresponding hole. The sorting of cards is achieved by passing a needle through the holes on the different cards at the particular position representing the desired classification. When the needle is raised the slotted cards fall off. Those cards are counted.

The *tally sheet* is a simple working sheets for making frequency tables. It has a tabular form with proper classification headings. As each data sheet (not sorted beforehand) is looked into, a tally mark is made at the appropriate classification cell in the tally sheet. The tally marks are later counted. To facilitate counting, the tally marks are blocked in groups of five.

In fully manual methods, while much of the arithmetical work will have to be done by direct computation, conversion tables and mathematical tables of squares, cubes, square roots, reciprocals and the like would be of considerable help.

5. *Mechanical aids*

Manual methods are often supplemented by mechanical aids. Thus, for instance, in counting operations that are likely to strain human memory a *count meter* may be used. For adding operations to obtain subtotals and totals of characteristics, *adding machines* are an aid. Some adding machines have large capacity and permit totals, together with the individual values, to be printed on paper tapes. These printed values can be compared with original sheets to check for correctness. The *abacus* may be used for all the four arithmetical operations and the *slide rule* for multiplication and division.

For the operations of addition and subtraction as well as for computations which would involve the other two operations of multiplication and division, desk calculating machines may be used. There are many types of calculating machines: multiple key board and simple key board types; the rotary and key actuated types; manually operated and electrically operated types. Most of these calculating machines permit the calculation of the sum, the sum of squares and the sum of pair-products of a large number of positive or negative values, as 'unit operations' (i.e., only the final result needs be copied down from the machine).

The *punched card system:* The punched card system which includes different types of mechanical equipment is a self-sufficient system for data processing. Before processing operations on punched card machines can begin, it is necessary to transfer the basic data to cards of a standard type by punching holes on the cards at pre-designed positions. There will be one card (or one set of cards) for data relating to each individual. Once the data are transferred to punched cards and checked for accuracy, the operations of duplicating copies, sorting or other arrangements, addition, subtraction, multiplication and division of values, and printing the results can be performed on the different machines in the system, at high speed. The system is suited to large scale work and offers the advantages of mechanization, such as high speed, savings in time, and facility to control accuracy of data.

The *electronic computer:* The electronic computer is a recent development. It can perform a long sequence of arithmetical operations much faster than most other machines. Instructions to the machine and the data are fed through punched cards, magnetic tapes or other media, and are stored in the 'memory' unit of the machine. The machine chooses the appropriate operation according to the instruction programmed in advance, from a set of alternatives given, and gives the results on magnetic tape, paper tape or punched cards.

6. *Control of quality*

Records such as output diaries, progress reports, and equipment utilization statements provide a basis for overall control of processing work. Apart from this, inspection of individual processing operations, is necessary. In some operations, such inspection can be done on the basis of built-in checks. In other cases, complete and independent duplication of the work will be necessary. If the degree of inaccuracies is expected to be small, checking the operation on a sampling basis or duplication of part of the work only will suffice.

In cases where the same set of data is processed under different classifications to give different sets of tables or other results, consistency between the different sets of results can be easily verified, especially when guided by the frequencies and subtotals and totals of counts or values.

Numerical accuracy of summary results should be checked by repeating the computation, when manually obtained.

Final results of a statistical investigation should whenever possible be compared with results of similar investigations and differences, if, any, should be explained.

In punched card methods or in using electronic computers, the emphasis should be on the programming of the job and the checking of pilot-run results; since, once the programming is correct and the machine is in order, errors can seldom occur.

Training on Topic 5.1

7. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 6 above.

(2) Consideration of leading principles in processing, such as accuracy, uniformity and economy and measures taken to achieve them.

(3) Detailed descriptions of items of work under different processing operations.

(4) Consideration of specimens of: (*a*) schemes and instructions for scrutiny of data sheets, (*b*) flow charts of processing operations, (*c*) working sheets for transferring information, (*d*) forms for summary results, (*e*) final results of processing.

8. *Demonstration*

Examples:

(1) *Scrutiny of figures.* Selected sets of figures which contain errors open to scrutiny will be used. Some of them may be on actual scrutinized data sheets. Why the figures are suspicious, how the errors could have arisen, how to tackle the errors and what will be the result of the scrutiny as regards the disposal of the sets of figures will be explained. Examples of figures selected may reveal the following: (*a*) wrong totals, (*b*) blanks left, (*c*) transcription errors, (*d*) wrong calculations, (*e*) mixing up of units, (*f*) wrong time reference, (*g*) counting and measurement errors, including rounding errors, (*h*) inconsistent entries, (*i*) illegible entries.

(2) *Punched card operations.* This demonstration will be supplemented by a visit to a punched card installation. The demonstration should be confined to a general introduction to the punched card system, with emphasis on the transfer of information from data sheets to punched cards. Specimens of punched cards, code lists, card designs, and lists, tables and other results printed on machines should be presented. Detailed description of the card itself, how coded information is transferred to the card by punching and how the punching is verified should be explained in detail. How sorting and other arrangement on sorters and collators and how the arithmetical operations and printing results on tabulators, calculators and other special machines are performed should be explained. The speed of the machines, the time saved, the accuracy possible and the cost involved should be briefly indicated. Mention should also be made of numerical work on electronic computers.

9. *Exercises*

Examples:

(1) Scrutiny of a number of filled-in data sheets of the same type, with the aid of standardized instructions. A copy of the instructions and a set of about 20 completed, but unscrutinized, data sheets should be made available to each trainee. Preparation of note based on the result of scrutiny.

(2) Coding a number of scrutinized data sheets of the same type. The coding should suit transfer to punched cards and involve quantitizing attribute forms with the aid of given code lists and making up uniformity in number of digits of variate forms. Preparation of note about sufficiency of code list, or about imperfections noted in data sheets. One set of code lists and instructions for coding and about 20 data sheets should be made available to each trainee.

10. *Visits*

(1) Data processing division of a statistical office where data are processed by manual operations. An office engaged in compilation of routine statistics may be given preference.

(2) Data processing division of a statistical office which uses a punched card machine installation. (Alternatively a demonstration at a punched card machine company office may be arranged).

11. *Training material*

Lecture notes. Exercise sheets. Material for exercises (section 9 above). Material for demonstrations (section 8 above). Other material for lecture and discussion (section 7 above).

TOPIC 5.2

Computation techniques

1. *Rounding*

In making observations, measurements have always to be rounded; counts may also be rounded. Moreover, because of the limited digit capacity of computing equipment, numbers will have to be rounded off suitably. Rounding a number to n digits is done as follows: first, discard all the digits to the right of the n-th digit, and replace digits, if any, discarded from the left of the decimal point with zeros; next, if the (n + 1)th digit that was discarded was less than 5 leave the n-th digit unaltered, if it was more than 5, or was 5 followed by at least one non-zero digit in some position, raise the n-th digit by one unit, but if it was just 5, leave the n-th digit unaltered if the n-th digit is even and raise it by one unit if it is odd. Thus the numbers 1632.36, 428723, 61.8501, 37.25, 42.15 and 5.103, rounded to three digits will be 1630, 429000, 61.9, 37.2, 42.2 and 5.10 respectively.

(The rule favouring an even digit to be left unaltered if the discarded digit is just 5 is only conventional. The purpose of such a rule is to avoid accumulation of rounding errors in summation.)

2. *Significant digits*

If a measurement is recorded as 143.3 mms., it is to be expected that the correct length lies between 143.25 mms. and 43.35 mms. Then the number of significant digits in the measurement 143.3 mms. is four. A digit to be called significant, requires that the error in the next position to its right is not more than 5 units in that position. The number 537852 may be rounded to three significant digits as 538000. But if a number 63200 is given without mention about any rounding made, it will be difficult to say whether the number of significant digits is three, four, five or six. To avoid this difficult, it may be presented as 6320×10^2 (or 6.320×10^5) if it was the case that there were four significant digits in the number. Zero will not be a significant digit when it is used to fix the decimal point or to fill the places of discarded digits. Thus in 0.0035020 the number of significant digits is five, it being assumed that the last zero has been retained as a result of rounding. The number of significant digits in a number gives an idea of the number's accuracy. Thus in 375×10^3 it is evident that the accuracy of the number is only within limits of 500 units on either side of 375000 or within $\frac{500}{375000} \times 100 = 0.13$ per cent limits.

3. *Error involved in operations with approximate numbers*

(1) In a sum or a difference, the absolute value of the error will not be greater than the sum of absolute values of the errors of the components.

Thus in $15.2 - 17.6 + 18.3 - 12.8$ where each number is rounded to the first place of decimal, the maximum error possible is $.05 \times 4 = 0.2$.

(2) In a product or quotient, the value of the relative error will not generally be greater than the sum of the relative errors of the components.

Thus in 18.23×1.264 the relative error will not be greater than $(\frac{0.005}{18.23} \times 100 + \frac{0.0005}{1.264} \times 100)$ per cent.

The result of calculation should not contain digits which are not significant. However, it is not so easy to find out the correct number of digits to be retained when a result is obtained by performing many different operations on numbers which are rounded and not exact. The rules given below may be followed when the numbers included in the operation are only a few.

(1) In addition and subtraction, more decimal places than there are in the original number with the fewest significant decimal places should not be given in the result.

Thus $21.62 + 15 = 37$
$835.1 - 16.21 = 818.9$

If a given number of significant decimal places is required in the sum or difference, then, in the component numbers, one more significant decimal place should be retained.

When a correct total is separately known, it should be used instead of the total obtained from components which are rounded.

(2) In multiplication, division or extraction of square root, more digits than there are in the original number with the fewest significant digits should not be given in the result.

Thus $184 \times 12 \times 367 = 81 \times 10^4$

$15.7 \div 16 = 0.98$

$\sqrt{0.0256} = 0.160$

If a given number of significant digits is desired in the result, each of the original figures should at least have that number of significant digits.

4. *Computational errors*

Apart from errors due to rounding or other approximations knowingly made, errors during computation can occur in various ways. Mistakes are very often made at the time of reading values from a source such as a document, a mathematical table or the counter of a calculating machine and copying or transferring the values. The usual mistakes of this category are substitution of one digit for another such as 0 for 6 and 9, 1 for 7 and 4, 3 for 8 and 7 for 9. Another type of mistake is the interchanging of digits, such as 43 for 34, or the copying of 14445 as 1445 or the omission of one or more 53's from the series, 50, 51, 52, 53, 53, 53, 54 . . .

To avoid such mistakes, the rules given below may be followed.

(1) Write figures distinctly.

(2) Cut down reading and transferring operations to a minimum.

(3) Use well-designed computation lay-outs.

(4) Provide for checks such as those based on mathematical identities.

(5) Anticipate the magnitudes of results beforehand.

(6) Reduce fair copying from rough work, to a minimum.

(7) Plan computational procedures beforehand.

(8) Break up computational steps into unit operations.

(9) Refer to mathematical tables for as many figures together as possible at a time.

(10) Follow instructions for the use of mathematical tables.

(11) When mistakes are detected, go over the computation all over again using a different computation lay-out, if possible.

When the procedure does not provide for internal checks, the only reliable means of checking and ensuring the accuracy of results is to have the computation independently repeated.

5. *Aids to computation*

Among the aids to computation are instruments like the abacus and slide rule, desk calculating machines, punched card machines and electronic computers, each type having its role depending on the volume of data and complexity of calculation (Topic 5.1).

Of equal importance are mathematical tables, giving values of reciprocals, squares, cubes and higher powers, square roots, cube roots, logarithms, exponentials, sines, cosines, tangents and values of other mathematical functions. So also are statistical tables providing values such as areas, ordinates and percentage points of probability distributions.

Training on Topic 5.2

6. *Lectures and discussion*

(1) Presentation of contents of sections 1 to 5 above.

(2) Brief review of computation of summary figures covered under Topic 1.6 and Topic 1.7 bringing out differences due to use of desk calculating machines.

(3) Introduction to methods and formulae for obtaining with the help of desk calculators: means, standard deviations and pooled means and pooled standard deviations from: (*a*) ungrouped series, (*b*) frequency tables, with a view to enable trainees to work out the exercises of types (1) to (6) below.

(4) For those at the *intermediate level* only — Introduction to methods and formulae for: (*a*) construction of index numbers, (*b*) standard errors of summary figures, (*c*) correlation and regression from ungrouped series and two-way frequency tables, (*d*) fitting linear trend, (*e*) interpolation and (*f*) solution of equations — with a view to enable trainees to work out the exercises of the types (7) to (13) below.

7. *Demonstration*

Example:

Use of desk calculators and other computational aids. The first part of the demonstration will be the use of manually operated desk calculators. Each trainee as well as the instructor will have a machine for use during demonstration. The position of the machine, and the manner of holding and operating the handle should be first indicated. The functional parts and capabilities of the machine should be explained. Addition and subtraction, and then multiplication and division as repeated addition and subtraction respectively, should be explained. Special features of the machine such as possibilities of obtaining sums, sums of squares, sums of products and even sums of quotients (division being equivalent to multiplication by reciprocal), as 'unit operations' should be pointed out. Time-saving features and short-cut methods and checks which it is possible to carry out on the machine should be indicated. These may be done by working out examples, the trainees following the instructor in using the machines.

The second part of the demonstration should be concerned with the use of mathematical tables separately, and in conjunction with machines, for computation. Examples of taking square roots, logarithms, reciprocals from the tables should be provided.

8. *Exercises*

The exercises proposed below are on numerical computation, using manually operated calculating machines, mathematical tables and other computational aids. Exercises 1 to 6 are for all primary and intermediate level personnel, 7 and 8 (marked with asterisk) for all intermediate level personnel, 9 to 13 (marked with two asterisks) only for those intermediate level personnel who are concerned with analytic work. Mathematical symbols, computation formulae, and leading principles will be covered during lectures and discussions (section 6 above) while hints, checks and lay-outs may be indicated on the exercise sheets.

Examples:

(1) Practice in: (*a*) rounding off numbers to specified number of: (*i*) decimal places, (*ii*) significant figures;

(*b*) calculating the errors in approximation: (*i*) absolute errors, (*ii*) relative errors;

(*c*) nature of combination of errors in arithmetical operations: (*i*) addition and subtraction and (*ii*) multiplication and division.

(2) Practice in the handling of desk calculators and mathematical tables. The possibility of obtaining, on desk calculators, the sum of squares or sum of products of positive and negative values, methods of handling

of varying number of decimal places in figures and of using mathematical tables in conjunction with machine work (example: multiplication with reciprocal instead of division) should be brought out.

(3) Computation of mean and standard deviation from an ungrouped series. The series should be long enough containing about 100 values.

(4) (*a*) Computation of ratios, percentages and rates from frequency tables.

(*b*) Combining several ratios, percentages and rates based on unequal numbers.

(5) (*a*) Computation of mean and standard deviation from frequency tables.

(*b*) Computation of mean and standard deviation from frequency tables, with change of origin and/or scale of the values of the variable.

(6) Pooling means and standard deviations based on unequal numbers of observations.

*(7) Computation of index numbers using different formulae.

(*a*) as means of ratios;

(*b*) as ratios of weighted totals.

*(8) Computation of standard errors using given formulae.

(*a*) standard error of a sample proportion;

(*b*) standard error of mean in sample from a finite population;

(*c*) standard error of mean in sample from an infinite population;

(*d*) standard error of standard deviation in sample from an infinite population (normal).

**(9) Computation of the first four moments and measures of skewness and kurtosis.

(*a*) from ungrouped series;

(*b*) from frequency table.

**(10) Computation from ungrouped bivariate data

(*a*) regression straight line;

(*b*) correlation coefficient;

(*c*) standard error of estimate in regression;

(*d*) standard error of correlation coefficient.

**(11) Computation from a two-way frequency table
  (*a*) regression straight line;
  (*b*) correlation coefficient.

(Change of scale and change of origin of the variables to be introduced)

**(12) (*a*) Fitting a straight line trend to time series data;
  (*b*) Graphic and linear interpolation;
  (*c*) Interpolation using Lagrange's formula.

**(13) Solution of linear, quadratic and simultaneous equations.

9. *Training material*

Lecture notes. Exercise sheets.

Course 6

PRESENTATION AND PUBLICATION OF STATISTICS

This course is intended for trainees at the primary as well as the intermediate levels. The purpose of the course is to provide understanding of the broad principles and acquaintance with the details of different methods of presenation of statistics and of publication of statistics.

Training will be given through lectures, discussions, demonstrations and exercises. To increase effectiveness, part of the training on each topic may be organized in the context of activities of a selected section of a statistical office engaged in the type of work falling under the topic.

Course summary:

Topic	*Number of hours proposed for instruction*			
	Lecture/ Discussion	*Demonstration*	*Exercise*	*Total*
6.1 Tabular presentation	6	2	6	14
6.2 Graphic presentation	6	2	6	14
6.3 Textual presentation and publication	6	–	4	10
TOTAL:	18	4	16	38

TOPIC 6.1

Tabular presentation

1. The systematic presentation of statistical results in their numerical form, in rows and columns designated by classification headings, is known as tabulation. Statistical data will have to undergo the processing operations of arrangement and condensation before they can be presented in tables. Some of the results of processing data on punched card tabulators, or by manual methods, may have already taken the shape of primary statistical tables. These tables also will often have to go through some refinements before they can be presented.

2. *Purpose*

One of the main purposes served by tabulation is to render the data handy without losing relevant information. This is achieved during processing. By tabulation, qualitative data can be rendered into quantitative form. Tables make the location of information easier. They enable assessments, comparisons and other interpretations to be made at a visual level. Tabular presentation can throw more light on statistics and may even suggest the nature of analysis that may be carried out. Tables provide a very convenient means of presenting results of large-scale work and are invariably used to present the results of routine compilation.

3. *Parts of a table*

The different parts of a statistical table are the following:—

(1) Table number. This is the identification number and precedes the title of the table.

(2) Title. The title is placed above the main body of the table. It gives, in brief and concise language, information such as: what kind of data are shown in the table, how the data are classified, where the data apply, and to what period the data apply.

(3) Caption or columnar headings: The 'caption' consists of the designations at the top of the column in a table explaining what each column represents. Each column designation is called a 'caption heading'.

(4) Stub or row designations: The stub is the left-most column or set of columns of the table, providing the designations for each row in the table. It will be noted that the stub column also bears a caption heading. Each row designation is called a stub item.

(5) Body of the table: The body of the table consists of the numerical information which are placed in appropriate cells governed by column and row headings.

(6) Source note: If the information presented has not been collected originally, it is required that the source of the data be given. Occasionally, a prefatory note and one or more foot-notes may be appended to a table.

(7) A prefatory note provides an explanation concerning the entire table or a substantial part of it. It is placed just below the title and in smaller or less prominent type.

(8) Foot-notes provide explanations concerning individual figures, or a column or row of figures. It is placed at the bottom of the table and also in smaller or less prominent type as the prefatory note.

4. *Types of tables*

Tables may be classified according to various criteria. A table the body of which consists of frequencies only is called a frequency table. Tables in which the classification items appear only on the caption side or only on the stub side are called one-way tables; others are called two-way tables. If the classification along the stub or caption is further sub-classified, the table is called a multiple classification table. A one-way table with multiple classification is called a hierarchical table. Tables produced by the punched card tabulator are hierarchical tables. On the electronic statistical machine, one can produce among other types of tables, two-way multiple classification frequency tables.

5. *Construction of tables*

A table should be expressive by itself and facilitate reference and interpretations. A general-purpose table would give full details that all possible users are likely to be interested in. A special purpose table would place emphasis on information relevant to a purpose and may omit unnecessary information. Every table should be complete within itself; i.e., it should contain explanations to make the meaning of each entry clear. The different parts of a table should be so arranged as to: (1) lay emphasis on important items, (2) make comparisons easy, (3) stress the important comparisons, (4) economize space and (5) have an attractive appearance.

In order to build in quality in a statistical table, attention should be paid to the following: (1) considering each part of the table separately, (2) choice and arrangement of stub and caption items, (3) numbering the columns and rows, (4) providing margins, spaces, bold letters, double lines and other devices to bring to easy notice the contents and prominent features, (5) providing subtotals and totals wherever desirable, (6) mentioning source of and limitations in the results presented and (7) overall neatness.

To construct a table, the classification characteristics will first have to be chosen from within the statistical material available for tabulation. Some of these may be qualitative and the others quantitative characteristics.

When there are more than one classification characteristic, they may be allotted to the stub as well as the caption and major and sub-classification characteristics will be positioned in an appropriate order. The different classes under each characteristic are then decided upon. These classes may be arranged in an alphabetical or a customary order when the classification characteristic is qualitative, and in order of magnitude when the characteristic is quantitative. An example of qualitative classification is geographical divisions arranged in alphabetical order. An example of quantitative classification is a time series in which the years are arranged in numerical order.

The items including sub-totals and totals which will go into the body of the table will be determined by the choice of the classification characteristics. The positioning of classification headings therefore should keep in view the need for making comparisons and for providing for those sub-totals and totals which are most desirable.

One of the first factors to be considered in the construction of a table is its lay-out. The lay-out is decided upon by considering the size and shape of the table and its direction and position on the page. A rough draft may be prepared first. The lay-out may be analyzed in terms of: (1) width of stub, (2) width of each column, (3) space needed for title, prefatory note, foot-note and source note, (4) number of lines needed for caption, (5) number of rows in body, (6) type of vertical and horizontal ruling, (7) extent of margins and spacing and (8) size and style of lettering.

6. *Principles in preparing tables*

When more than one table is prepared, they should be serially numbered. For table numbers, the Arabic decimal system is suitable especially when there are many groups of tables.

The title should be clearly worded and brief but adequately indicative of the contents of the table. If necessary, the title may be supplemented by a prefatory note giving an explanation of the table.

The captions or the prefatory note should mention the units in which the figures are expressed. If the figures in the table are rounded, notes should be provided indicating the degree of approximation made.

Totals, sub-totals, averages and percentages should be provided in the table whenever desirable. Percentages or averages may preferably be presented along with the numbers on which they are based.

In presenting results of sample surveys, the number of sample units associated with each set of results should be given, and as necessary, standard errors or coefficients of variation of estimates, also presented.

When no cases have been found to exist or the value of an item is zero, the fact is indicated by dots or short dashes and footnotes provided. It is not customary, especially in general purpose tables, to show a zero alone.

The caption should be briefly worded and should cover designations of units. Two or more columns under one class may be grouped under a 'spanner-heading'.

If a stub item is very long it should be written in more than one line. No lines are drawn in the stub between major and sub-classification items.

Explanations concerning individual figures or a column or row of figures should be given in foot-notes with suitable keys for identification.

A source note may appear below the title or prefatory note or below the foot-note. The source should be completely identificable in terms of author, title, volume, page, publisher and date.

With regard to ruling and spacing, tables are not usually closed at the sides. When closed they should be closed on both sides. No horizontal lines are used in the body of the table except to demarcate sections and totals.

Too much variety in types should be avoided. But separate types may be used to distinguish foot-notes, and other notes from the body of the table.

Training on Topic 6.1

7. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 6 above.

(2) Explanation of the different parts of a table and details in the preparation of tables, with the help of specimen tables.

(3) Presentation of tables of different types and with differences in detail.

8. *Demonstration*

Example:

Remodelling defective tables. This demonstration may be made in either of two alternative ways. Different selected tables each deficient in one or more respects, or a single table in which several defects are artificially created may be used. The latter will be easier and more effective in a brief demonstration. A table which has an unconventional table number, a long and vaguely worded title, incorrect stub and caption allocations, mixed reference to units, no column numbers, no proper ruling, no sub-totals, ineffective spacing, lack of uniformity in number of decimal places, and many other defects should be presented to start with. Each defect should be pointed out and each detail attended to one by one and a revised table evolved on the black-board during the demonstration.

9. *Exercises*

Examples:

(1) Preparation of blank tabular forms most suited to presentation of specified statistical results. Attention to parts of the table and other details of presentation are to be emphasized. The types of tables that may be covered are: non-frequency tables, hierarchical tables and multiple classification tables.

(2) Rendering of textual presentation of statistical results into statistical tables. The textual matter selected should be such as to reveal the effectiveness of tabular presentation in respect of the matter.

(3) Remodelling of deficient tables. Actual instances of tables inefficiently presented should be selected, as far as possible. Different small tables each containing one or more defects will have to be selected, as it will be seldom possible to find a single table with many defects.

10. *Training material*

Lecture notes. Exercise sheets. Specimen tables for explanation of parts, details and types. Specimen defective table(s) for remodelling during demonstration.

TOPIC 6.2

Graphic presentation

1. The rendering of statistical results in the form of charts is known as graphic presentation. Charts are geometrical representations of magnitudes. Some charts adopt geometrical forms such as rectangles and segments of circle which vary in dimensions (usually length or area) or pictorial forms which vary in the number of times they are repeated in proportion to the magnitudes represented. Some other charts use the magnitudes as co-ordinates to generate points, line curves or surfaces.

2. *Purpose*

Charts are intended to give a quick overall impression of the information contained in statistical results. They are more effective than tables in attracting attention and are more easily understood. Charts help to bring out relationships that may be hidden in tables. However, charts are suitable for showing only brief sets of statistical results, whereas tables are capable of presenting very extensive statistical series. The values shown by a chart are likely to be more approximate than those that are given in a table. For these reasons, charts often accompany tables, or numerical values are sometimes in-set on charts. Charting requires a certain amount of artistic ability.

3. *Types of charts*

Some of the charts commonly used for presenting statistics are the following:

(1) Pictogram: This chart uses equal-sized pictorial forms. Thus the picture of a cement bag of specified size, may represent a hundred thousand bags of cement.

(2) Pie chart: This is a representation for comparing magnitudes of component parts of a total. The component parts are shown as sectors of a pie or a circle.

(3) Bar chart: In bar charts, magnitudes are represented by size of rectangular bars. There are several types of bar charts to suit different kinds of statistics. Some of the types are: the simple bar chart, multi-bar chart, plus-minus bar chart, component bar chart, shaded and over-lapping bar chart, and pyramid bar chart.

(4) Line chart: In this type of chart, magnitudes are plotted to give a line curve so as to bring out characteristics such as trends, patterns and changes. Some of the common types of line charts are: simple line graph, component line graph, balance graph, high-low graph.

(5) Statistical map: This type of chart is used to present geographical statistics. Among the different types of statistical maps are: (1) cross-hatched or shaded maps, (2) dot maps, (3) pin maps, (4) combination maps.

4. *General requisites*

A well-constructed chart should satisfy the following requirements: (1) neat general appearance, (2) well-proportioned lay-out, (3) suitable lettering, (4) brief but explanatory title, (5) source reference, (6) comparability of statistics presented.

5. *Construction of bar charts*

Bar charts are specially suited to represent statistics which are classified according to a qualitative criterion, although occasionally they may be used to represent time series or frequency tables. The bars referring to the same characteristic on a bar chart should be of the same width and vary only in length. Bars may be placed either horizontally or vertically. The bars should be arranged in some systematic order applied to the classification criterion or to the length of the bar. Thus they may be arranged according to the alphabetic order of names of geographical regions or in decreasing order of length of bar. Usually the space between two bars will be less than the width of a bar. In a histogram there will be no space between adjacent bars.

The bars may be coloured, shaded or outlined and the classification clearly written out against each bar, preferably horizontally. Scales may be provided on the chart to facilitate reading of magnitudes. The zero base line or other base of comparison should be indicated. In multiple bar charts or component bar charts, bars or components relating to the same characteristic should be shaded alike and legends provided.

6. *Construction of line curves*

The line curve is ordinarily drawn only when the classification characteristic is a variable. Some of the basic steps in drawing line curves are the assignment of variables to different co-ordinate axes, choice of scale and base line, and paying attention to details regarding the plotting of points and drawing the curves.

The curve representing the statistical magnitudes is generally ruled more heavily than other lines appearing on a chart. When different curves are drawn on the same chart, curves are made distinguishable using solid, dotted and dashed lines or by heavy and light lines or by lines of different colours. In component line graphs, suitable shading will be necessary. All lettering on a chart including scale labels, scale values, legends or curve designations and any other words or figures should be placed at appropriate places, preferably horizontally.

7. *Constructing statistical maps*

In a cross-hatched or shaded statistical map, a progressive change in the shading will indicate an increase (or decrease) in the characteristic. In a large-dot map the variation in magnitudes is indicated by the number of equal-sized large dots in each geographical area, the number of units represented by a dot being large enough so that the number of dots apparing in each area is small. In a small-dot map, the number of units represented by a dot is small enough and located at the exact place where it is applicable so that the many dots can give the effect of gradual changes in intensity of shading. Pin maps are mounted on backing of cork or board on which values are represented by means of pins with heads of different sizes, colours and shapes. Pin maps may be readily altered as facts change. A statistical map in which the magnitudes are represented by bars instead of dots is an example of a combination map.

8. *Some other charts and their uses*

Some other types of charts which are of analytic value rather than devices for final presentation of results are the following:—

(1) The simple line graph used for the purpose of interpolating the value of one variable corresponding to a given value of another variable.

(2) The scatter diagram useful in detecting relationship between two variables. Association between the variables will be indicated by a clustering of points on the scatter diagram round a sloping trend upward or downward.

(3) The semi-logarithmic graph, a convenient representation when relative changes and not absolute changes in magnitudes are of interest.

(4) A control chart, essentially with a central line round which values plotted are expected to fluctuate, and upper and lower control limit lines which almost no value should cross unless the characteristic is influenced by assignable causes, or in other words unless there is lack of control.

(5) The histogram, frequency polygon and frequency curve useful in obtaining relative frequencies of variate values.

(6) The ogive useful in graphically obtaining fractiles (median, quartile, decile, percentile and so on) of a variable characteristic.

(7) The Lorenz curve, a useful representation for comparing patterns such as those of income distribution or commodity consumption and to measure departures from the line of equal distribution.

Training on Topic 6.2

9. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 8 above.

(2) Explanation of details of different charts, with the help of specimens.

(3) Explanation of: (*a*) choice of type of chart to suit the statistics to be presented, (*b*) choice of scale for charting (arithmetic scale, logarithmic scale, broken scale, scale captioning) and (*c*) charting techniques (principles of charting, drawing and colouring instruments and materials).

10. *Demonstrations*

Example:

Charts—Special features and purposes served. Specimens or wall-charts of a variety of graphic presentations will be necessary for this demonstration. Special features in charting pictograms, pie charts, multiple bar charts, component line charts, balance charts and similar other charts should be pointed out. Besides, the purposes served by different types of charts should be explained on the lines indicated in sections 2, 3 and 9 above.

The instruments and material commonly used for charting should be exhibited and the methods of using them explained. This demonstration may be held in the charting section of a statistical office or may be supplemented by a visit to such a section.

11. *Exercises*

Examples:

(1) Construction of pie charts and simple multiple bar charts of different types, from given statistics.

(2) Charting line graphs, component line graphs, balance graphs and other line curves based on given statistics.

(3) Preparation of statistical maps, and other special charts from given statistics.

12. *Training material*

Lecture notes, Exercise sheets. Wall-charts of important types of graphical presentation. Special charts of a wide variety, preferably based on national statistics.

TOPIC 6.3

Textual presentation and publication

1. *Textual presentation*

The presentation of statistical results in textual form is done by means of explanatory notes and technical notes attached to statistical results presented in another form, or by means of a comprehensive report. Explanatory notes and technical notes are usually brief and are used for purposes of giving background information of statistical results or for presenting comments and conclusions about information contained in statistical tables and charts. Some technical notes may deal with matters such as sampling design. Explanatory notes and technical notes contain such matter as would form part of a statistical report if a full report were written.

2. *Report writing*

Full reports are usually written for special investigations rather than for presenting routine statistics. A reason for this is that routine statistics are released periodically and at close intervals so that, if a report were written each time, much of the textual matter would have to be repeated. Secondly, routine statistics being mostly general purpose statistics, it will be an almost impossible task to provide a full report giving analyses and interpretations that would serve the variety of possible uses of the statistics.

The quality of a report will very much depend on the technical and writing abilities of those responsible for producing it. However, a report on a statistical investigation should provide descriptions of the various administrative and technical aspects of the investigation, besides presenting the results of analysis along with interpretations and conclusions.

The report on a statistical investigation may contain chapters or sections on the following: (1) Introduction, giving the background, and purpose. (2) Design, explaining the coverage, the concepts and definitions used, and details of sampling techniques, if any, adopted. (3) Data collection, covering method adopted and details of collection. (4) Cost, giving an account of the organisational arrangement and cost incurred on personnel and equipment and for the different operations. (5) Accuracy and limitations of the data. (6) Results of analysis and inference, consisting of tables, charts, summary results and technical notes on the results. (7) Conclusion. (8) Appendices giving other information, such as specimens of data sheets and instructions for investigation.

A draft report is usually made first and then circulated for comments, before the final report is made. Examples of reports are census reports and sample survey reports.

5. *Statistical publications*

Statistical publications are the end results of statistical activities. A report or a periodic bulletin when published, carries with it the mark of authenticity and reliability. Prompt and timely publication of reliable statistics should, therefore, be the ultimate goal of all statistical systems.

Statistical publications may be classified broadly into four groups:

(1) General publications issued regularly (monthly, quarterly, or annually), such as monthly bulletins of statistics, or statistical yearbooks.

(2) Special subject publications issued regularly (monthly, quarterly, or annually), such as publications on external trade statistics, agriculture and food statistics, vital statistics, industrial statistics, transport statistics.

(3) Census publications which appear, usually, once every ten years or sometimes every five years. Very often, preliminary results for the country as a whole are first published in brief volumes, to be followed later by a series of volumes giving detailed information.

(4) Special or *ad hoc* survey reports issued according to demand.

6. *Routine publications*

The format of a routine statistical publication and the pattern of its publication are mainly determined by factors such as: (1) the time required to produce the publication, (2) the budget available, (3) the amount of statistics that will be put into the publication and (4) the possible uses of the statistics that will be published.

The timing of a publication—that is, determining the periodicity of issue of the publication, whether it is to be weekly, monthly, bi-monthly, quarterly, half yearly, annual or bi-ennial — is an important problem. Publications should contain current figures. On the other hand, a certain amount of time will be required to collect data, to process them and to have the results printed and distributed. General publications such as bulletins and year books will have to be confined to summary statistics for various reasons. A certain amount of time will have to be allowed before statistics from various sources are received and can be combined and summarized.

To maintain the periodicity of a publication, efforts will have to be made to avoid bottlenecks and to minimize the time interval between data collection and printing results. However, results should not be released for printing unless they satisfy certain quality standards. Bottlenecks frequently occur at the printing stage. This may be due to lack of good printing facilities, due to awaiting parts of the publication to be got ready or due to making corrections at the last moment. Even after printing, delays can occur before the publications are distributed for use.

Routine publications should maintain their continuity and uniformity of coverage. Firstly, therefore, concepts and definitions and classifications should be uniform and standardization is unavoidable. Even if the reference period of a series is not of the same duration in each issue of the publication, continuity of the series will have to be maintained to keep the series complete.

Since a number of publications with different periodicity will be issued, proper co-ordination of the publications at all levels, will be necessary to avoid duplication and to ensure that a balanced and up to date set of information are available at the national level.

Since demands for statistics change depending upon changes in economic and social conditions, it will be necessary to review publication programmes periodically. Such a review may be made keeping in view the following:

(1) elimination of unnecessary publications, (2) ensuring that each publication retained meets a need, (3) improvements in the content and format of the publications, (4) considering whether a new publication deserves to be introduced or not, (5) standardizing publications wherever possible.

5. *Details of publication*

Several organizational arrangements are necessary for the preparation of publication material. When a draft of a publication is prepared, a committee of experts may examine the text, tables and charts and help to finalize the publication.

In the preparation of publication material, details will have to be attended to in order to ensure that the tables, charts, computed results, descriptive notes, technical notes and conclusions are prepared and arranged in the most useful manner possible. Attention will have to be paid to style of writing, choice of titles and numbering titles, paragraphs, tables and charts. Arrangements for proof correction, publication of errata, and distribution and sale will also be necessary. All these details should be covered in the publication plans and programmes that may be prepared in advance.

Care should be given to the preparation of statistics in publications. Base year periods, units, definitions and limitations in the data should be carefully explained in explanatory notes or in introductory pages. Very often statistical offices refrain from repeating concepts and definitions in every publication. In such cases, explicit reference should be made to the publication where descriptions of concepts and methods used can be found.

It is important to publish statistical data in a form that can be easily understood by the readers. Too many details should be avoided. For example, when publishing data on foreign trade, it is seldom necessary to publish the figures in units smaller than one million national currency units. Similarly, population figures will, usually, be expressed in 000's. The date to which the statistics refer should always be clearly indicated. For example, population figures may refer to 1 January or 1 July. Agricultural statistics may refer to agricultural years. Fiscal years may run over a twelve month period different from calendar years. The beginning and end of such years should be indicated by a foot-note. For example: "Data refer to fiscal years beginning 1 April of the year stated".

Difficulty arises when the method of compilation of a series has been changed, so that figures after the change are not strictly comparable with figures before the change. The user may be warned against such changes in methods, by suitable notes or other indication. Sometimes provisional figures requiring revision may have to be published in tables. Clear indications of provisional figures and revised figures will have to be made.

A statistical publication usually covers a small number of years only. Care should be taken that all series are shown for the same set of years, because many users may wish to relate statistics on one subject to those on other subjects.

The scale in a chart should neither be so compressed as to [illegible] all fluctuations, nor so [illegible] as to exaggerate the fluctuations in relation to the fluctuations shown in other charts of the same publication. It is sometimes found that publications do not contain tables showing the statistics on which the charts are based, or the tables differ from the charts. Sometimes, the period covered by a statistical chart is different from that covered by the corresponding table. Errors in units, particularly when logarithmic scales are used, may easily occur.

Any conclusion in the text of a publication should be substantiated by statistics. Hence the text should not deviate from the tables. It is not desirable to draw conclusions which the reader cannot verify from the statistics published.

Articles of a mathematical type may require the printing of mathematical symbols and formulae. It has sometimes been found useful to include all mathematical formulae in appendices. After the publication of results of a statistical investigation, it may be found helpful to prepare a more popular statement for the newspapers usually in co-operation with newspapermen. The results when published in newspapers should be watched for accuracy.

As statistical publications are used by international organizations, universities, government agencies, and business enterprises all over the world, many countries find it useful to publish English or French translations of the headings of tables and explanatory notes, or to append a summary in English or French.

Training on Topic 6.3

6. *Lectures and discussions*

(1) Presentation of the contents of sections 1 to 5 above.

(2) Consideration of a list of items of work in the publication section of a selected statistical office.

(3) Familiarization with selected national statistical reports and routine publications, in respect of the publication programmes behind them, their format and get-up.

(If possible, (2) and (3) should be carried out, in the context of activities in the publication section of a statistical office, so as to ensure contact by trainees with actual situations.)

7. *Exercises*

Examples:

(1) (*a*) Organization of manuscript for publication by consolidating given tables, charts and textual matter.

(*b*) Correction of proof-sheets (received from printer).

(*c*) Preparation of errata (after printing).

(2) (*a*) Rearrangement of a jumbled table of contents of a statistical report.

(*b*) Writing descriptive notes based on given tables and charts.

8. *Training material*

Lecture notes. Exercise sheets. Copies of specimen list of items of work in the publication section of a statistical office. Specimen statistical reports and periodical publications.

Course 7

UTILIZATION OF STATISTICS

Training at the primary level need be concerned with only the first of the two topics under this course. Trainees at the intermediate level will go through the entire course. The objective is to provide acquaintance with some of the general and analytic problems involved in the use of available statistics for specified purposes.

Many of the prerequisites to utilization are covered by other courses. Thus, a general knowledge of the statistical series available within and outside the country and the broad purposes for which they may be used are dealt with under Topic 1.2 and also Topic 9.1. Considerations of factors affecting the quality of statistics that may be used are given in Topic 1.9. Computational techniques for any reduction of statistics that may be necessary prior to utilization are covered in Topic 5.2 mostly through the exercises under that topic.

Training will be given through lectures, discussions and exercises. No separate demonstration is provided for; but during lectures, exercises and discussions, case studies of utilization may be exhibited.

Course summary:

Topic	*Number of hours proposed for instruction*		
	Lecture/Discussion	*Exercise*	*Total*
7.1 Some guides to utilization	4	2	6
7.2 Some methods in utilization (for intermediate level only)	(6)	(8)	(14)
TOTAL:	4 (10)	2 (10)	6 (20)

(The figures in brackets show hours to be spent by intermediate level personnel only)

TOPIC 7.1

Some guides to utilization

1. Although published statistics carry with them a reasonable guarantee of reliability, no user can afford to accept them as a base for his decisions and actions without knowing their meaning and limitations. Secondly, he should be able to ascertain the adequacy of the statistics in respect of the purpose in hand. Apart from this, the user should know the methods of utilizing them in the most efficient manner with respect to his problem or purpose. Utilizing statistics in an efficient manner involves a knowledge of the different statistical tools for analysis and interpretation of quantitative information, whether based on sampling or not, and of the different pitfalls that are likely.

2. *Some basic rules*

A number of general rules may be given, which do not involve any analytic method, but will provide guidance in making inferences and decisions directly on the basis of the statistics available. Firstly, it is necessary to consider carefully what statistics ought to be collected for the purpose in hand, what sources should exist and what are the existing sources of information. The existing sources are then examined for sufficiency as regards the problem in hand. When the sources of information have been selected, the next step is to understand the exact definitions of the entities and characteristics involved in the statistics. Thus the question may be asked: What is a registered birth? or What is personal income?

After the selection of statistics concerning some characteristics, the next question is how are they related to the characteristics actually wanted; how good are they as measures?

In dealing with summary figures such as averages, percentages and rates, one should examine whether the standards of reference used are appropriate for the purpose in hand. Thus a possible question is: Has the birth-rate been reckoned per 1000 of the population or per 1000 married couples?

Before comparisons are made between magnitudes, their comparability should be ascertained. Accurate comparisons can be made only between closely similar entities or over quite short periods.

Another point to be verified is the degree of accuracy of the statistics. Are the results accurate enough for the conclusions based on them to be valid? Conclusions about a recurring phenomenon based on isolated results should not be relied upon, since temporary fluctuations are likely to be mistaken for permanent changes in such cases.

Having determined the exact meaning and limitations of the statistics, consideration should be given as to the conclusions they can lead to, or as to whether they are inadequate so that no valid conclusions can be based on them without further investigation or else without additional statistics to fill up gaps. Adjustments of the available statistics may have to be made to suit the purpose in view, such as by filling gaps through estimation, or by reclassification.

Sometimes, statistics available from one source will not agree fully with those for the same purpose from another source. Methods of reconciling them, wherever feasible, and removing any obvious discrepancies may have to be adopted for proper utilization. Also, inconsistencies between results of sample surveys and statistics of the by-product type may have to be resolved. When there is no regular method for deciding such cases, those statistics which appear to be more accurate or reliable according to the judgment of the user may have to be accepted.

3. *Pitfalls to be avoided*

Statistics may be handled differently by different users, depending on the experience, skill and objectivity they possess. Statistics may be used in the wrong way and for wrong purposes. Interpretations which are not valid may be made on their basis. Some of the pitfalls to be avoided in using statistics are given below:—

(1) *Changing definitions:* Shifting definitions in a statistical series would make the items non-comparable and, if comparisons are made of magnitudes based on different definitions, the comparisons will not be valid. Examples of shifting definitions are: personal income, which might be salaries and wages or income from all sources; and city population, which might be population of corporation area or population of corporation area and suburbs.

(2) *Misleading expressions:* Either falsely to support a claim or unintentionally, invalid or exaggerated statements may be made, quoting statistics. Thus one may claim that, during a five-year plan period, 66 per cent of the hydro-electric projects were put into action (whereas there were 3 projects only in the plan). If the number of workers leaving with less than 3 months service falls from 500 to 300, it may only mean that fewer workers were engaged in one quarter than in the preceding one. Another type of misleading expression is contained in the following examples: the average man was 28.9 years old in 1933, he is 35.3 in 1963, in 30 years he has aged only by 6.4 years; men were buying 0.2 blankets per year.

(3) *Technical errors:* Due to imperfect knowledge or negligence, technical errors in data processing or presentation may be committed. Examples are omission of weights in calculating an average of averages or an average of percentages, or errors in units of measurement or confusion between natural and common logarithms. Charts with scales magnified unduly and use of wrong formulae are further examples.

(4) *Inappropriate comparisons:* Comparisons of magnitudes based on different definitions may even lead to opposite conclusions. Thus the price of a commodity during the current week may indicate a fall compared to the previous week and a rise compared to the previous week last year. The conclusion that average earnings went down by 8 per cent whereas every worker was receiving something more would seem absurd, if the reason that higher paid positions were discontinued were not taken into account.

(5) *Misinterpretation of relationship:* The fact may be ignored that statistical evidence of relationship does not necessarily mean causal relationship. Thus it will be wrong to conclude that the price of corn is dependent on fever cases on the basis of a statistical correlation, or to conclude that, due to population increasing, the death-rate decreases.

(6) *Disregard of variability:* The average of a value without regard to the variability in the values which make the average, can lead to erroneous conclusions. Thus it may be wrong to conclude that a yield of 50 units per hectare is an exceptionally high yield because the average yield is only 45 units. It will be wrong to point out as an inconsistency the fact that in two regions people have average incomes nearly same, but one has many poor people.

(7) *Wrong generalization:* Wrong interpretations may be made by generalizing results obtained from isolated cases or biased samples.

(8) *Ignoring sampling error:* Estimates based on samples may be identified with population characteristics, ignoring given measures of sampling error.

(9) Statistics may be taken out of their context and used to establish a case or point. Deliberate dishonest uses are not to be ruled out. Thus to establish that the incidence of a disease was very high, the number of visits to hospital by patients, including many visits by the same patient in connection with one attack, may have been quoted, instead of the number of persons attacked.

Training on Topic 7.1

A. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 3 above.

(2) Examples of selection and examination for adequacy, of statistics for different specified purposes. Examples of adjustments made to make statistics adequate and usable.

(3) Examples of pitfalls of different types.

5. *Exercise*

Example: Selection of statistics for stated purposes, from statistical publications and examination as to their adequacy.

6. *Training material*

Lecture notes. Exercise sheet. Copies of statistical publications. Specimen case studies in which statistics have been selected from several sources to meet different purposes.

TOPIC 7.2

Some methods in utilization

(for intermediate level only)

1. In utilizing statistics, those selected may be directly used or they may have to be subjected to further analysis before they can be interpreted and serve the purpose required. The totals given in a population census report, the estimates in a sample survey report, the totals of economic magnitudes in a governmental or international publication, and published index numbers are examples of statistics that may be used directly without further analysis. But if a life-table is to be constructed, if a special set of indicators is to be calculated, if the trend of production of a commodity over several years is to be determined, if a comparison of consumption patterns is to be made, if association or correlation between two characteristics is to be measured, if the joint effect of a number of related characteristics is to be ascertained, if estimates based on sampling are to be compared, or if an assumption or hypothesis is to be verified, the statistics available for the purpose may have to be subjected to further analysis and interpretation. A number of methods are available for analysis and interpretation in these and other situations. The methods of analysis, which may be graphical or numerical, may be broadly classified as those which are not based on sampling principles and those which have to take into account fluctuations due to sampling.

Some of the methods are briefly described below. Computational techniques concerning these have mostly been covered already in Topic 5.2 (exercises). The principles involved and methods of interpretation are to be given emphasis under this topic.

2. *Smoothing a series*

Since it will be possible in practice only to obtain a few of the values of a characteristic intended to measure a phenomenon, it will be necessary to smooth out concentrations due to biases or other erratic situations in the values, in order to arrive at the general law of behaviour of the characteristic. This may be done in several ways: (1) by plotting the values and drawing a free-hand smooth curve, (2) by computing moving averages of appropriate length, (3) by fitting a mathematical function and graduating the values according to that function.

3. *Fitting a trend*

The trend in a time series is the general pattern of change of the magnitude in question, over a long interval of time. It may be determined by: (1) drawing a free-hand curve over a plotting of the values, (2) calculating moving averages, (3) fitting a mathematical function, by a method such as the method of 'least squares'.

4. *Construction of index numbers*

Index numbers are a general class of indicators used for comparisons over time, between localities and so on. Index numbers may be obtained as ratios of individual values, of simple sums of values, or of weighted sums of values, or may be arrived at as simple or weighted arithmetic or geometric means of relatives. Thus a consumer price index may be computed as a ratio of the sum of the prices of different consumption items in one month and the sum of the prices of the same items in another month used as the base for comparison; the two sums of prices may be suitably weighted by the respective quantities of consumption; the statistics relevant for the purpose being retail price statistics based on market prices and consumption pattern statistics based on family budget surveys. Some of the common formulae used for the construction of index numbers are Laspeyre's, Paasche's and Fisher's formulae.

An important class of indicators used in the analysis of time series are index numbers of seasonal variation. They may be obtained by a graphical method, method of moving averages or some other method.

5. *Concentration curves*

The concentration curve, also known as the Lorenz curve is an effective graphical means to represent and compare distribution patterns in respect of characteristics such as income or commodity consumption. Suppose that the data consist of the expenditure on food items by the different persons in a community. These data can be rearranged in two columns with expenditure on food items presented in ascending order of magnitude and the number of persons shown against each expenditure value. Two new columns are worked out giving successively cumulated values corresponding to the original columns. These cumulated values are further expressed as percentages of the total expenditure and total number of persons respectively. The concentration curve will be obtained on plotting the results by representing the cumulative percentage of persons on the horizontal axis and the cumulative percentage of expenditure on the vertical axis. In this representation, a line drawn diagonally through the origin will represent the line of equal distribution and the departure of the concentration curve from this line will be an indication of the disparity in expenditure pattern among the persons in the community.

A variation of this curve is possible for representing two characteristics instead of one. Thus on one axis the cumulative percentage of persons ordered according to expenditure on food, and on the other axis the cumulative percentage of expenditure on sugar may be represented. A concentration curve so obtained will indicate for instance, the percentage of amount spent on sugar by those persons who belong to the lower 25 per cent class with respect to expenditure on food items.

6. *Association, correlation and regression*

Relationships that may exist between two attribute characteristics are technically referred to as association and those between two variables as correlation.

There are several methods of detecting association between two attributes. These are based on a study of the two-way frequency table (called contingency table) that will be prepared from two-attribute data. One of the summary figures used for measuring and detecting association is the coefficient of contingency.

The study of correlation between two variables may be made from bivariate data in their ungrouped form, or grouped into a two-way frequency table. The scatter diagram of the data is used to detect the existence of correlation. The extent of correlation is measured by calculating a coefficient of correlation. Another problem is the determination of the law of relationship between two variables so as to enable the prediction of one variable, given a value of the other variable. For this purpose, the regression equation is determined on the basis of bivariate data.

When one variable can be viewed as a joint effect of several other variables, the method of regression of one variable on another, is extended to regression of one variable on a set of other variables. This is known as multiple regression. Thus a multiple regression equation to predict the export of textiles on the basis of production of jute, production of cotton, domestic consumption of textiles and the value of gold may be determined using multivariate data on these five characteristics.

7. *Methods allowing for sampling*

When the statistics have been based on sampling, necessary allowance will have to be made for sampling fluctuations while making interpretations. Many of the methods dealing with sample statistics consist of calculating measures of sampling errors and using probability distributions.

Thus the reliability of sample estimates of population characteristics may be judged with the help of the standard errors of the estimates. A sample mean may be compared with a population mean, making allowance for sampling error with the help of the standard error of the mean. Two sample means may be compared taking into account sampling errors in both the means by using the standard error of the difference between the two means. The association between two attributes may be tested from a sample contingency table, by evaluating the probability of a suitable test criterion called the 'chi square'. A correlation coefficient calculated from sample bivariate data may not indicate real correlation in the population; it will have to be judged along with measures of the sampling error in the correlation coefficient, which in turn depends on the sample size.

Training on Topic 7.2

8. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 7 above.

(2) Illustration of such techniques for analysis mentioned in sections 2 to 7 as are not covered by exercises under Topics 5.2 and 6.2.

(3) Illustration of possible uses of the results obtained by analysis mentioned in sections 2 to 7.

(4) Interpretation based on sample estimates, with the help of their standard errors.

(5) Interpretation based on sample criteria for testing hypotheses.

9. *Exercises*

Examples: (Simple exercises covering variety to be given, involving preparation of notes based on interpretation).

(1) (a) Interpretation of statistics presented in tables.

(b) Interpretation from charts (semi-logarithmic chart, concentration curves, scatter diagrams).

(2) (a) Construction of index numbers from given series.

(b) Calculation of seasonal indices after removal of trend component.

(3) (a) Chi square test for association in contingency tables.

(b) Standard error test for proportion and means.

(4) (a) Tests for correlation coefficient and regression.

(b) Simple exact tests and analysis of variance for one-way classification data.

10. *Training material*

Lecture notes. Exercise sheets.

Course 8

SUPERVISION OF STATISTICAL WORK

This course is for trainees at the *intermediate level* only. It is intended to acquaint personnel who are to carry out supervisory work in statistical offices with the functions and requirements of a supervisor and to provide knowledge about details and methods of efficient supervision. Of the two topics under this course, the first is supervision in general while the other is supervision of operations such as field work and data processing.

Training will be given through lectures, discussions, exercises and visits. No class room demonstration is provided for. But during lectures and discussions, specimen forms and charts used for purposes of supervision will be exhibited. An important method of training in this course will be demonstration during visits to selected places where supervisory work of different types are carreid out. If the trainee participates in project work (appendix 1), he will have the opportunity to observe how the teaching staff carry out supervisory work of different types.

Course summary:

Topic	*Number of hours proposed for instruction*			
	Lecture/ Discussion	*Exercise*	*Visit*	*Total*
8.1 Supervision in general	4	2	2	8
8.2 Supervision of operations	4	4	4	12
TOTAL:	8	6	6	20

TOPIC 8.1

Supervision in general

1. Statistical operations such as data collection, processing and presentation involve the services of a number of primary personnel. To co-ordinate and exercise control over the work of the primary personnel, proper supervision is necessary. To be able to supervise effectively, the supervisor himself should know the job of the primary worker thoroughly and in addition have administrative and technical ability of a higher order. Some supervisors will be required to perform a variety of statistical duties while others may be specialized in particular statistical operations such as field work or processing.

2. *General requisites*

The supervisor should have the necessary education, experience, and temperament suited to the type of work he will have to perform. In addition to knowing the objectives, principles, procedures and other technical aspects of the job that he is to supervise, he should have an intimate knowledge of the subject field of the job. Thus a supervisor who is to look after field work in an agricultural survey should not only have technical knowledge about the statistical techniques involved, but also be well acquainted with local conditions, agricultural practices and factors affecting agriculture. The supervisor will have to maintain a high sense of responsibility in order to instil a similar sense in his subordinates and to help them adapt themselves to the tasks assigned to them. Supervisory staff, like the primary staff, will have to adhere strictly to the procedures and instructions set. However, the supervisor should not insist on standards of performance that are unreasonable. He should be able to recognize quickly outstanding performance and provide incentives to encourage it. A good supervisor will be understanding and tactful in his dealings with subordinates but firm in his decisions on matters requiring corrective action. He should closely observe the method of work of the primary staff and, when it is faulty or imperfect, be able to demonstrate how to do it correctly.

3. *Administration*

The supervisor's primary administrative function is to formulate and direct the activities of the staff under him. To do this efficiently he will have to maintain effective control and co-ordination of the different items of work.

The supervisor, whether he works in an office or in the field, will also have to perform many other routine functions since he has to handle staff, material and equipment and at times, money. He should be able to draft and issue written directives, instructions and memoranda to subordinate staff and correspond through letters with various types of persons, organizations and authorities, on matters concerning the job he supervises. He should be capable of preparing organization charts, designing forms for progress reports and other control data, and writing brief reports about his job.

4. *Planning and regulation of work*

Perhaps the most important part of the supervisor's work is the planning or scheduling of the job he is to supervise and implementing the plan or schedule. For this, he will have to resolve the job into a number of elements or operations, estimate the personnel, time and budget requirements, and determine the flow of work and anticipate end results. He should also incorporate effective checks on the quality of work. The implementation of the plan will involve instructions to the workers regarding the method of doing the work and their responsibilities in doing it, allocation of work items to different workers, handing over of instruction sheets and material for work, visiting workers at the spot of work and giving necessary advice, making arrangements for taking stock of work completed, checking the work for quality and finally putting in efforts to rectify imperfections and to reduce gaps in work and time lags, if any.

5. *Control*

To have jobs executed according to a plan, the supervisor will have to exercise control in various directions. For this, 'observation' on the work itself is the best method. Maintaining an integrated and complete set of records about various aspects of jobs, would facilitate supervision work. Some of the records that may be maintained are: inventory of work material (data sheets etc.), work allocation registers, workers' output diaries, machine log-books and staff movement control record (in field work). Such records would not only be an aid in routine control during execution of the work, but also will provide basic information for preparing progress reports, setting performance standards, preparing budgets, assessment of wastage, preparing incentive schemes and various other purposes.

6. *Training*

The supervisor, apart from being required to come to the rescue of the primary worker whenever needed, should be able to impart training, individually or in groups, for enhancing efficiency of the workers. He should be able to prepare detailed instruction manuals on the operations and demonstrate systematic methods of carrying out the operations, where necessary, using instruments and other equipment and pointing out pitfalls and precautions.

Training on Topic 8.1

7. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 6 above.

(2) Consideration of specimen lists of supervisor's duties, selected from suitable statistical offices.

(3) Consideration of specimen forms and charts maintained for supervisory work, selected suitably.

8. *Exercise*

Example: (*a*) Preparation of form for workers' diary, for the following types of routine work:

(*i*) Scrutiny of data
(*ii*) Computation

(*b*) Preparation of form for register of work allocation to processing staff. (The background and specification of each type of work will be furnished. The information, when collected on the forms, should be useful for purposes of control.)

9. *Visit*

A chosen section, under the control of a supervisor in a statistical office, with a view to demonstrate how a supervisor plans, co-ordinates, directs and controls the work under his charge. If possible, arrangements existing for training workers should also be demonstrated.

10. *Training material*

Lecture notes. Exercise sheet. Specimen lists of supervisor's duties and specimen forms and charts aiding supervision.

TOPIC 8.2

Supervision of operations

1. *Field work*

In carrying out data collection operations by field work, ordinarily, each supervisor will be responsible for the work done by a group of field workers. The supervisor himself may have to move from the headquarters and maintain a camp office. The supervisor's responsibility will be firstly to ensure that the workers have fully understood the instructions for data collection and that they will be able to do this in full accordance with the requirements placed on the data to be collected. To statisfy himself on this aspect he may make each worker complete a few experimental data sheets by field work. He should be able to provide instructions, clarify doubts and demonstrate correct methods, by accompanying each worker to the spot of observation in the initial stages. Secondly, he should make sure that they are fully equipped with copies of data sheets, instruments, note books, writing boards and other material and equipment.

The next function of the supervisor will be to assign specific amounts of work that each worker will have to complete in a specified time period. Field workers may be asked to report to him each day, if possible, to hand over completed data sheets, to present any problems confronted and to seek clarification.

Data sheets returned by field workers should be checked by the supervisor as soon as received, and any incomplete or wrong entries should be refererd back immediately for rectification. When it is not possible for a supervisor to check all the data sheets returned to him, he may check them on a sampling basis, paying more attention to those workers who are likely to produce relatively poor quality material. So also, a check on the physical movements of field workers, guided by the movement control sheet that he may maintain could be made on a sampling pattern.

The supervisor will have to attend to problems such as providing substitute field staff in case of absence due to ill health, or otherwise, of workers. He will have to advise workers on matters such as modes of travel and time suitable for contacting respondents. He may also have to interpret and decide on matters on which the written instructions are silent.

It will be the duty of the field supervisor to despatch to the head office completed and checked data sheets according to the instructions, without any delay.

Different supervisors engaged in an investigation may consult one another on problems of common interest. The supervisor should keep in touch with the head office by visits, or by sending periodic reports if he is far away.

An important function of the supervisor will be to collect a part of the data independently by himself and thus place a check on the reliability of the data collected by his workers. In some cases, as in a crop cutting experiment, the supervisor may have to be present at the spot of observation.

2. *Processing*

In processing work, the supervisor will be concerned with one or more of the following items of work: (1) receiving, storing and handling of data-sheets, (2) scrutiny and editing of data sheets including coding operation, (3) transfer of data to punched cards or to working sheets, (4) computation. In these, the supervisor is expected to have the operations carried out by the primary workers by exercising control over the accuracy of results, by setting up fool-proof and uniform methods and achieving economy in terms of time, personnel and resources.

It is the function of the supervisor to classify the range of processing operations into the most advantageous divisions and determine the sequence of the different stages of work. For this, he should know the quality and quantity of work that can be expected in a specified time, for different alternative procedures that may be possible. He should be able to decide which items of work it will be economical to perform manually and which mechanically. So also he should be able to allocate each item of work to the right type of worker and regulate the flow of work, avoiding bottlenecks and idle hours.

His own experience and past performance data will enable him to prescribe standard output or time rates for the different elements of each operation. These he will use to evaluate jobs, to prepare time estimates and judge worker and machine performance. Progress records of individual operations and informative charts will be maintained wherever possible for routine control.

In computational work, the supervisor should be able to prepare the most efficient lay-outs, prescribing short-cut methods where possible and providing built-in checks. He should be capable of detecting mistakes quickly and as a routine, conduct sample checks on numerical work.

The supervisor will also have to carry out a continuous appraisal of workers, techniques, equipment, and other material from the point of view of maintaining a high level of efficiency.

He should consult and work in unison with co-supervisors in the same office who handle similar or allied operations to ensure uniformity of procedures and results. Material processed from one section under a supervisor will often have to pass to another section for a further stage of processing. Co-operation and consultation between the supervisors of the clearing section and receiving section are also essential for smooth flow of work. (The above apply equally whether processing operations are manual or by punched card methods.)

3. *Presentation and publication*

The preparation of final tables and charts, out of processed data, for presentation purposes also require supervision, especially in large-scale work involving a number of workers, or punched card machines. What counts most in a supervisor looking after such work is his technical knowledge and craftsmanship and ability to obtain uniform results of high quality from different workers, which when put together will give a set of effective tables and charts. A supervisor of charting should himself be a good draftsman.

Supervision also assumes importance in offices concerned with the publication of statistics. Co-ordination and control of activities such as preparation of manuscripts for the press, proof correction, preparation of errata and distribution of the publications will be necessary.

4. *Compilation processes*

When statistics are compiled from records which arise as by-products of administration, and the compilation work is undertaken by the same office which produces the records, the supervisor of compilation will have to perform a variety of functions. This is so because such offices often do not have full-fledged statistical sections with technical experts in different aspects of statistical work. The supervisor will then have the additional responsibility of the technician. The functions of planning, executing and reviewing the compilation work will be his responsibility. A versatile and competent supervisor will therefore be necessary to look into the various aspects of compilation work. A supervisor in such a position may have to secure advice on technical matters from technicians in statistical offices.

He should have good background knowledge of the material used for compilation, be capable of preparing compilation schemes including data sheets, and directing the operations of data extraction, processing, presentation and publication (sections 2 and 3 above).

Training on Topic 8.2

5. *Lectures and discussions*

(1) Presentation of contents of sections 1 to 4 above.

(2) Detailed examination of supervisory methods in: (*a*) fieldwork, (*b*) processing work, (*c*) presentation and publication work.

(3) Detailed examination of supervisory methods special to a compilation office.

6. *Exercise*

Examples:

(1) Preparation of the following for a specified investigation:
  (*i*) field work programmes
  (*ii*) field workers' dairy form
  (*iii*) field workers' movement control sheet.

(2) Preparation of the following for a specified investigation:
  (*i*) flow-chart for processing work.
  (*ii*) progress chart concerning an operation.
  (*iii*) budget estimates for a specified job.

7. *Visits*

(1) The section under a supervisor dealing with field work in a statistical organization, with a view to demonstrating supervisory functions and methods relating to field work. In addition, if possible a spot where the field work is carried out in the presence of a supervisor, as for example in a crop cutting experiment.

(2) The sections under supervisors, engaged in: (*a*) data processing and (*b*) presentation and publication, in a statistical office, with a view to demonstrating supervisory work concerning those activities. An office which has a separate charting section should be chosen, if possible. The statistical office selected may be one engaged in large-scale compilation work, preferably using punched card machines.

8. *Training material*

Lecture notes. Exercise sheets. Specimens of forms used for controlling field work, flow charts used in data processing, cost estimates for jobs.

Course 9

STATISTICAL ORGANIZATION, CO-ORDINATION AND STANDARDS

This course is for trainees at the *intermediate level* only. The main purpose of the course is to provide knowledge about the statistical system of the country, the net-work of agencies furnishing the various types of statistics, how their work is co-ordinated to form an integrated system and what important statistical series are produced. Trainees should, in particular, know the role that their own statistical offices play in this system.

In order to provide the necessary background, the following also will be briefly covered: functions of a national statistical system, types of statistical systems, types of statistical offices, national statistical co-ordination, co-ordination procedures, statistical standards, legal provisions and acquaintance with international statistical activities.

Training will be given with the help of illustrations drawn from national activities and through lectures, discussions, exercises and visits. No separate demonstration other than those during visits is provided for; but during the training, attention will be drawn to material showing the structure of the national statistical system and the working of statistical offices, as also to national statistical publications and series.

Course summary:

Topic		*Number of hours proposed for instruction*			
		Lecture/ Discussion	*Exercise*	*Visit*	*Total*
9.1	Statistical systems	6	2	2	10
9.2	Statistical co-ordination and standards	6	2	2	10
	TOTAL:	12	4	4	20

TOPIC 9.1

Statistical systems

1. The importance of statistics and the demand for them in the economic and social activities of governments have led to the growth of national statistical services. Different countries may have different types of statistical systems. A national statistical system covers all official statistical activities in the country and includes organizations such as central statistical offices, specialized statistical offices in ministries or departments, statistical offices in the political sub-divisions of the country and even official statistical committees or boards.

2. *Functions of a national statistical system*

The main function of a national statistical system is regularly to collect, compile and publish an integrated set of statistics relating to the economic and social structure and activities of the country. The first task of the system is therefore to determine what statistics are required and how best to produce them. In this task, a periodic review of the statistics being produced will be necessary. Traditional statistical series may have to be improved, supplemented and mutual connections between the series established. This will require research and special studies. To develop integrated statistics, the national system should have a general plan which will have to be prepared in consultation with the various users of statistics. The plan should aim at co-ordination of the different statistical agencies in the country by recommending common definitions, classifications and techniques, proper coverage and comparable time references. Gaps in current statistics and the priorities with which they are to be filled should be indicated. The plan should ensure that information is not unnecessarily duplicated by different agencies and that money is not spent for statistics which do not meet a definite current or prospective need.

Among the other functions of a national system are: (1) Development and issue of statistical standards adapted to the country's needs. These will consist of standard definitions, standard classifications and standard statistical methods. (2) Research in statistical operations and techniques. (3) Co-ordination of statistical activity of different national offices. (4) Training of personnel. (5) Maintaining quality, including integrity, independence and objectivity of statistics. (6) Participation in international statistical activity.

3. *Types of statistical systems*

Different types of national statistical systems are in existence.

(1) A decentralised system of various national statistical offices each specializing in a subject field, control and co-ordination being at a minimum. A committee may only be responsible for overall national statistics.

(2) A system decentralized by subject, with a co-ordinating authority. This is the system prevailing in some large countries.

(3) A centralized system with a major operating office for general statistics and specialized offices under a co-ordinating authority. Co-ordination of activities of specialized offices in individual departments, may be the responsibility of the central office or another co-ordinating authority.

(4) A fully centralized system consisting of one central statistical office responsible for all statistics. Such a system has the advantage of locating technical knowledge and equipment in one place making co-ordination a simpler process. To protect the objectivity and independence of the statistics produced, a national council may be established with control over the central office.

4. *Types of statistical offices*

The different statistical agencies or offices in a country may be classified as follows:

(1) Large-scale field organizations such as those for censuses and sample surveys.

(2) Large-scale data processing organizations.

(3) Agencies concerned with statistics arising as by-products of administration.

(4) Specialized statistical organizations such as statistical offices in departments of agriculture, and industry.

(5) Central co-ordinating offices.

(6) Central statistical offices with operating responsibilities.

Training on Topic 9.1

5. *Lectures and discussions*

(1) Presentation of the contents of sections 1 to 4 above.

(2) Providing a descriptive account of the statistical system of the country, indicating the type, structure and the various functions performed.

(3) Providing a list of the statistical offices and other institutions in the country, giving descriptions of their organization and functions, the statistics produced by them and the publications issued.

(4) Familiarization with the important statistical series of the country.

(5) Presentation of brief descriptions about the activities and publications of international statistical offices, such as those of the United Nations and the Specialized Agencies.

6. *Exercise*

Example: Writing a description of the statistical office to which the trainee belongs, indicating its place in the national system, organizational set-up, nature of work, and results produced; preparation of organization chart of the statistical office.

7. *Visit*

A statistical office producing statistics on a large scale, with a view to understand the organization of the office, the different divisions and sections in the office and the flow of work within the office.

8. *Training material*

Lecture Notes. Exercise sheet. List of statistical offices and other institutions in the country. List of important national statistical series. List of important international statistical series. Organization charts of selected statistical offices.

TOPIC 9.2

Statistical co-ordination and standards

1. The statistical system in a country may be either centralized or decentralized depending on the degree of overall control and co-ordination existing in the activities. Co-ordination is necesasry for obtaining an integrated system possessing efficiency as well as uniformity in methods and results. Statistical standards are a means for achieving uniformity in statistical results.

2. *National statistical co-ordination*

In most countries, especially those with a decentralized statistical system, a number of more or less independent agencies will be concerned with the production of statistics. Co-ordination of the activities of those agencies is necessary to minimize inefficient utilization of technical knowledge, duplication of work, gaps in information and excessive budgets and to ensure that the most useful set of statistics is produced, at the national level. National statistical co-ordination would be concerned with co-ordination of both the statistical services in the country as well as the statistics produced by them. Special legislation may be required to give the co-ordinating agency sufficient power to operate and to secure the co-operation of the different statistical offices. The type of co-ordination needed would vary according to the structure of the statistical system prevailing in the country. The nature of the system will determine whether co-ordination is to be based on legislation or internal directives, whether there should be a council of departmental representatives or a committee, whether co-ordination should be effected on a legal basis or through voluntary co-operatiion. Co-ordination will affect all producers of statistics in and outside the government, whatever be the type of statistics produced or kind of statistical operations performed by them, as also the users of statistics. Co-ordination plans should therefore pay attention to the difficulties and problems of the producers as also take into account the needs of the users.

3. *Co-ordination procedure*

For satisfactory co-ordination, the responsibility will have to be fixed on some one person or organization enjoying legal authority, autonomy, technical capacity, and the co-operation of all concerned. Co-ordination may be exercised through: (1) central co-ordinating office, (2) national statistical council, (3) co-ordination committee, (4) national statistical conferences.

The different measures for co-ordination will include: (1) planning work programmes, (2) reviewing data sheets, (3) establishing statistical standards, (4) selecting and exchanging personnel, (5) preparing and issuing co-ordinated statistical publications such as year-books.

The achievements of a co-ordinating agency require periodic review and can be judged from the following: (1) the nature of the plan for national statistics, (2) technical quality of methods, (3) extent of duplication of work and expenditure, (4) consistency and comparability of data, (5) extent to which users' requirements are met, (6) timeliness of release of statistics.

4. *Statistical standards*

Statistical standards are necessary for the maintenance of quality levels and comparability of statistics. Statistical standards differ in detail according to the subject fields to which they relate and the particular purpose the statistics are intended to serve. Broadly they fall into the following three groups: (1) Standard definitions, such as those of employment, household, and establishment, (2) Standard classifications such as those of economic activities, trade commodities and diseases, (3) Standard methodology, such as methods for constructing an index of industrial production and methods of presenting sampling errors of estimates.

National standards will have to be adapted to suit the country and may differ from international standards. However for international comparability, national statistics released for international use should conform to international standards.

Standardization being a new development, the situation in most cases is one of adopting standards where none exist. In this respect standardization will be a step towards improving the quality of statistics. The advantages will be more if standardization is not confined to individual fields, but extended to related fields. Thus it should be possible to use the same definition of a household, both in a survey of housing conditions and in a demographic survey; or the same commodity classifications in a survey of production and in a survey of distribution.

Standard classifications are usually accompanied by standard code figures. The adoption of standard classifications therefore obviates the need for constructing different code lists from time to time.

There are now available international standards in different subject fields. These are designed to fit the economies of most countries as they are evolved out of national standards. In standards of clasisfication the categories are major groups which are further sub-divided into minor groups. Thus the standards afford some flexibility for expansion and contraction of classification categories. The decimal system of numbering categories facilitates coding.

5. *Legal provisions*

The execution of statistical policies and the organization and development of statistical systems can be considerably strengthened by legal provisions for the collection and publication of statistics. Legal authority in this respect may take the form of acts of parliaments and legislatures, ordinances, statutory rules and regulations, decrees, or administrative directives and instructions; it may also arise out of international treaties or agreements. Such legal provisions make it obligatory on the part of individuals public and private enterprises or other bodies concerned, to furnish information and at the same time they protect the respondents from non-statistical uses of the information supplied by them. In most countries there are laws regulating statistical activities.

Training on Topic 9.2

6. *Lectures and discussions*

(1) Presentation of the contents of sections 1 to 5 above.

(2) A detailed review of the situation in the country with respect to: (*a*) measures taken for co-ordination, (*b*) statistical standards adopted and (*c*) legal provisions existing.

(3) Familiarization with: (*a*) publications containing national and international statistical standards and (*b*) national statistical acts and other legal provisions.

7. *Exercise*

Example:

(*a*) Attempts at definitions of concepts such as productivity, sickness, unemployment, family, dwelling unit, in specified contexts.

Compiling accepted definitions of these concepts from published standards.

(*b*) Preparation of code lists for a set of given characteristics to suit a specified investigation.

8. *Visit*

A statistical office engaged in co-ordinating the statistical work of other offices, with a view to gaining knowledge about statistical standards and about steps taken to co-ordinate statistical services and statistical publications.

9. *Training material*

Lecture Notes. Exercise Sheet. Copies of national and international standards, standard code lists and statistical acts.

Appendix 1. PROJECT WORK

1. The project work which would form an important part of the training is intended to provide an insight into the inter-relations of the different operations in a full piece of statistical activity. This part of the training need be given only to those who are undergoing full training either through Courses 1 to 7 or Courses 1 to 9. It should have full demonstration value and at the same time be capable of giving the trainee some experience. The project will have to be carried out as a whole and be directed by the teaching staff who will also have to take over such of its stages as are beyond the capability of the trainees. But each trainee should have the opportunity to participate in as many items of the project work as his technical and general ability permits and to observe other items of work being carried out by others, in order to obtain a complete understanding of the nature, sequence and inter-connections of the different operations.

The project chosen should be as representative as possible of the various aspects of statistical work. For example, it may involve two types of statistical individuals (e.g. household and person); the characteristics to be observed may be such that some of them require measurement, some counting, some personal estimation and some interviewing a respondent; the data sheets to be used may be a combination of schedule and questionnaire. From the point of view of both simplicity and providing experience in sampling, it is desirable that the project should be a sampling investigation. Experience in field work will in general be more valuable for the trainees than in collecting data through the mail or by abstracting from records, and therefore, the project could be one which involves field work. A family budget enquiry, a survey of housing conditions, an opinion survey, a survey of manufacturing establishments or a survey of labour conditions or a similar investigation would be very suitable. On the other hand, if the trainees in a particular session are mostly those concerned with routine compilation work only, a compilation project involving data collection from records or by mail enquiry may be adopted instead.

3. Considering the total time available, the project has to be a brief one, involving only about 25 data sheets per trainee. It is desirable that it be undertaken from almost the beginning of the training and made to proceed side by side with the different Courses. Except for data collection, if it requires field work, all items of work concerning the trainees can be done in the class-room, in short sessions. As regards items of work other than data collection (field work or others, the project work can be divided into 20 stages as detailed below, a period of approximately 2 hours being spent on each stage. The stages to be taken up with each of the different courses is shown in table 1 (page 4).

4. The processing of data collected should in any case be done initially by manual methods, using desk calculators. Punched card machines, if available, may be utilized only for incidental demonstration and to duplicate and check the work done by the trainees, unless special training in punching and machine operations is intended as part of the training.

5. Items of work which are beyond the trainees ability, such as sampling design and report writing, if any, should be taken over fully by the teaching staff. To guarantee the quality of work, they should also step in at different stages of work, for purposes such as remodelling and finalizing data sheets, instructions for data collection, and lay-outs of tables, charts and computation forms. While each trainee is required to participate in as many items of work as possible, small convenor groups of trainees, to assist the teaching staff may be formed in respect of different items of work, such as: 1) design including making data sheet and instructions. 2) scrutiny, 3) tabulation, 4) analysis, 5) control and progress, 6) charting and 7) report-writing.

6. The sessions corresponding to the twenty project stages are described below:—

(If the project undertaken is not one involving sampling or field work, sessions 5 and 6 will not be necessary. Besides, some other sessions may be reduced in duration).

(1) The announcement of the project work is made. Its purpose, an outline of work and the programme are indicated. The characteristics involved, the method of observation and other details are discussed in a general way without entering into technical details. Each trainee is asked to prepare a list of suitable characteristics and present the list at the next session together with his proposals with regard to the individual to be observed, the method of observation and problems anticipated in conducting the investigation.

(2) The list of characteristics and proposals submitted by the trainees are discussed and a choice made of the population characteristics, the characteristics for observation, the individuals, the methods of observation and the methods of data collection. The convenor group responsible for designing is asked to consult teaching staff and present at the next session a revised list of characteristics.

(3) The list of characteristics prepared (copies of which are made available to all trainees) is considered and discussion is continued in terms of the concepts, definitions and procedures to be settled before preparation of data sheet and instructions are attempted. Details of designing the data sheets and preparing instructions are reviewed. The convenor group concerned is asked to submit a draft of the data sheet and instructions.

(4) Copies of the data sheet and instructions prepared are made available to trainees, discussed and improvements suggested. (The convenor group is asked to prepare a revised data sheet and instructions and the teaching staff would take over the draft from the them for remodelling and making a few provisional copies.)

(5) The type of sampling to be adopted is suggested by the teaching staff and the sampling procedure is broadly explained. The question of choosing a suitable frame is discussed. A decision is made about the frame to be used and the method of making the frame available.

(6) Depending on the sampling procedure and data collection method the trainees either obtain or construct afresh, the necessary frame. The frame is checked for being complete and up-to-date. The sampling units are then selected according to the procedures specified, and the list of sampling units prepared. (If a list of the frame is not necessary in advance for data collection, this stage is to be omitted. On the other hand, if a frame is to be constructed afresh, more time should be allowed).

(7) Copies of the provisional data sheet and instructions for filling it are made available. The data collection procedures are discussed by considering the instructions, one by one, along with the corresponding item or question on the data sheet. The plans for a try-out of the data collection are discussed. (Each trainee studies the instructions thoroughly, obtains clarifications from the teaching staff and completes two copies of data sheet during the try-out, outside class hours.)

(8) Programmes and plans for data collection, for receiving and scrutiny of data sheets, preparation of tables and charts, analysis and report writing, if any, are discussed in a general way and the different groups responsible for these items are asked to prepare provisional programmes and plans for presentation at a future session.

(9) The results of the try-out are considered. Difficulties experienced during the try-out and imperfections noted in data sheet and instructions are reviewed. (Incorporation of changes found necessary and the finalization of data sheet and instructions, including printing, are taken over by the teaching staff.) Changes in data sheet affecting the provisional programmes and plans prepared in stage 8 are also noted during discussion by the respective groups.

(10) and (11) The different programmes and plans prepared by the group are discussed and accepted for finalization. The group responsible for control and progress is asked to prepare a survey

calendar, making provision for work already done. (Data collection commences. Data sheets are returned without delay to the scrutiny group when duly completed.)

(12) When about one-fourth of the total number of data sheets have accumulated, the scrutiny work is started and shared between all the trainees, one trainee scrutinizing data sheets filled in by another. A discussion based on the findings of the scrutiny done during the session takes place and any change in procedure found necessary is communicated to all concerned.

(13) Completing scrutiny of data when all data sheets have been returned. Rejected data sheets, if any, are returned for fresh collection of data. The data sheets are arranged in order.

(14) General review of data collection and scrutiny already completed. Further work to be done and plans for allocation of sorting, summarizing, computation and other work, are discussed.

(15), (16) and (17) Manual sorting and preparation of summary tables. (When this stage is completed, the data sheets may be released for punched card work, if processing is to be duplicated on punched card machines.)

(18) and (19) Computation and preparation of final tables and charts.

(20) Results of analysis are discussed and interpretation is made. The scheme for presentation of results is discussed and decided.

(The results of the project will be consolidated by the teaching staff and put in the form of tables and charts with notes, or of a full report copies of which will be made available to the trainees, when prepared.)

Appendix 2. EQUIPMENT AND TRAINING MATERIAL

(For training a batch of thirty persons)

1. In order to conduct the training efficiently, adequate training facilities by way of equipment and training material are essential. Facilities available in statistical offices and other statistical institutions in the country may be used as far as possible. However, considerable preparatory work will have to be done to provide the necessary training material and acquire equipment. The equipment and training material should be ready and reviewed well before the training begins.

2. The minimum equipment and training material requirements are indicated below. Additional items, which are desirable but not essential, are also shown. Equipment would involve capital expenditure, while

expenditure on items such as stationery would be of a recurring type. The list of training material may require revision from session to session. The estimates are given, on the basis of 30 trainees per session.

3. ACCOMMODATION

   1. Classrooms for lectures, demonstrations and laboratory work with black-board and other furniture and fittings.

4. EQUIPMENT

   1. *Minimum list*

		Number
1.	Manual calculators—(1 per trainee plus 5 for demonstration/substitution in the case of breakdown)—total:	35
2.	Electric Calculator—(for demonstration)	1
3.	Mathematical tables—of squares, cubes, square roots, reciprocals, logarithms, etc.	
	Example:	
(*i*)	Barlow's tables	35
and (*ii*)	Chambers's or Castle's (logarithm) tables	35
4.	Statistical Tables:	
	Examples:	
(*i*)	Fisher and Yates	3
(*ii*)	Biometrika tables	3
(*iii*)	Random number tables	35
5. (*a*)	Dice (20 faces, for generating random numbers)	35
(*b*)	Dice (6 faces)	35
(*c*)	Playing cards (for demonstration of probabilistic problems)—sets	6
6.	Drawing instrument sets (including pantographs (lettering sets), drafting machines, planimeter, drawing board and tracing table)	6
7.	Copying machine (for reproduction of charts, etc.)	1
8.	Printing (mimeo) machine	1
9.	Typewriting machine	1
10.	Counters	3
11.	Stopwatches	3
12.	Serial Numbering machine	2
13.	Slide projector	1
14.	Other equipment (such as weights and measures, balances, crop-cutting instruments for field work)—sets	6

2. *Additional equipment desirable*

15.	Abacus	6
16.	Slide rules	6
17.	Movie projector	1
18.	Tape recorder (for reproducing interviews, etc.)	1
19.	Stereographer	1
20.	Punched card machines unit (on rental basis, if not borrowed)	
(*i*)	punches and verifiers (7 + 5)	12
(*ii*)	Sorters	2
(*iii*)	Tabulator (total transferring type)	1

5. SUPPLIES — Stationery and other recurring items (per session)

   1. *Minimum list*

      1. Stencils
      2. Paper, (printing, copying, typing)
      3. Paper, (squared, graph)
      4. Ink, (printing, drawing, colour)
      5. Other items.

   2. *Additional items*

      1. Photographic films and material
      2. Recording tapes
      3. Punched cards

6. TRAINING MATERIAL*

   1. *Minimum list*

      1. Training manuals and lecture notes specially prepared to suit the country's needs, for different topics in the syllabus.
      2. Work-sheets for practical exercises (laboratory work as well as field work). These sheets should contain selected exercises and hints for working out the exercises, with provision of squared and graph papers to work out the exercises.

---

* The material required for training on each topic is shown separately, in the earlier pages, in the last section under each topic.

3. Syllabuses, training programme, reference lists, instruction sheets, information sheets, etc.
4. Demonstration materials, such as schedules and questionnaires, instructions for investigation, punched card designs, town and village maps, tabulation lay-outs and apparatuses, as necessary, for the different demonstrations.
5. Wall charts and slides for display and illustration, such as organization charts, survey calendar, types of graphical representation.
6. Statistical publications—Copies of statistical reports, handbooks and manuals of the United Nations, FAO, WHO and other organizations, elementary text books for reference (at least 2 of each)

2. *Additional items*
   7. Photographs and movie films
   8. Tapes (recorded)
   9. Aerial photographs.

7. NOTE

Adequate provision should be made for repair and maintenance of calculating and other machines.

Appendix 3. PUBLICATIONS FOR REFERENCE

Only a selected, and not an exhaustive, list of publications is given here. The list is to be made up-to-date from time to time. The publications are not to be prescribed as texts for study.

(National publications which are not given here are to be added to this list, as necessary).

1. Books

1.	Arkin, H. and Colton, R. R. ..	Statistical Methods
2.	Croxton, F. E. and Cowden, D. J.	Applied General Statistics
3.	Mounsey, J.	Introduction to Statistical Calculations
4.	Wallis, W. A. and Roberts, H. V.	Statistics, A New Approach
5.	U.S. Department of Commerce, Bureau of the Census	Manual of Tabular Presentation
6.	Schmid, C. F.	Handbook of Graphic Presentation
7.	Yates, F.	Sampling Methods for Censuses and Surveys

2. Statistical and mathematical tables

1.	Comrie, L. J. (editor)	Barlow's Tables
2.	Pryde, J. (editor)	Chambers's Seven Figure Mathematical Tables

3.	Tippett, L. H. C.	Random Sampling Numbers: (Tracts for Computers, No. 15)
4.	Fisher, R. A. and Yates, F.	Statistical Tables for Biological, Agricultural and Medical Research
5.	Pearson, E. S. and Hartley H. O. (editors)	Biometrika Tables

3. International publications on methods

1.	UN Statistical Office	Handbook of Statistical Organization, (Series F)
2.	UN Statistical Office	A Short Manual on Sampling (Series F)
3.	UN/FAO	Handbook on Data Processing Methods
4.	UN Statistical Office	Industrial Censuses and Related Enquiries (Series F)
5.	UN Statistical Office	Principles of Statistics of External Trade
6.	UNESCO	Manual of Education Statistics
7.	UN	Problems and Methods in the Gathering of Representative and Comparable Wholesale Prices
8.	UN Statistical Office	Concepts and Methods in Statistics of Distribution (Series F)
9.	UN	Index Numbers of Wholesale Prices
10.	ECAFE	A Handbook on Agriculture Census Techniques in Asian Countries
11.	UN Statistical Office	The Preparation of Sampling Survey Reports
12.	FAO	Estimation of Crop Yields
13.	FAO	Methods of Collecting Current Agricultural Statistics
14.	FAO	Sampling Methods and Censuses, Part I, Collecting Data and Tabulation
15.	ILO	Methods of Family Living Studies
16.	ILO	Methods of Compiling Statistics of Industrial Disputes
17.	UN Statistical Office	Handbook of Population Census Methods (Series F)
18.	UN Statistical Office	Handbook of Vital Statistical Methods (Series F)

(For other publications, refer UN Statistical Office, Directory of International Standards for Statistics (including a bibliography on methods)).

4. International standards

1.	UN Statistical Office	International Standard Industrial Classification of all Economic Activities
2.	UN Statistical Office	International Standards in Basic Industrial Statistics
3.	UN Statistical Office	International Standard Definitions of Transport Statistics
4.	UN Statistical Office	Standard International Trade Classification
5.	WHO	Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death
6.	ILO	International Standard Classification of Occupations
7.	UN/FAO	World Weights and Measures: Handbook for Statisticians

(For other publications, refer UN Statistical Office, Directory of International Standards for Statistics).

5. Publications containing international statistics

UN Statistical Office

1. Statistical Yearbook
2. Demographic Yearbook
3. Yearbook of National Accounts Statistics
4. Yearbook of International Trade Statistics
5. Monthly Bulletin of Statistics

ILO

6. The International Labour Review—Monthly Statistical Supplement
7. The Yearbook of Labour Statistics
8. Occupational Safety and Health

FAO

9. Yearbook of Food and Agriculture Statistics: Part I—Production, Part II—Trade.

IMF

10. Annual Report
11. International Financial Statistics (monthly)
12. Balance of Payments Yearbook

WHO

13. Annual Epidemiological and Vital Statistics
14. Epidemiological and Vital Statistics Report (Monthly)

ECAFE

15. Economic Bulletin for Asia and the Far East (Quarterly)
16. Economic Survey of Asia and the Far East (Annual)

WHERE TO BUY UNITED NATIONS PUBLICATIONS AND THE PUBLICATIONS OF THE INTERNATIONAL COURT OF JUSTICE

AFRICA

CAMEROON:
LIBRAIRIE DU PEUPLE AFRICAIN
La Gérante, B. P. 1197, Yaoundé.
DIFFUSION INTERNATIONALE CAMEROUNAISE
DU LIVRE ET DE LA PRESSE, Sangmelima.

CONGO (Léopoldville):
INSTITUT POLITIQUE CONGOLAIS
B. P. 2307, Léopoldville.

ETHIOPIA: INTERNATIONAL PRESS AGENCY
P. O. Box 120, Addis Ababa.

GHANA: UNIVERSITY BOOKSHOP
University College of Ghana, Legon, Accra.

KENYA: THE E.S.A. BOOKSHOP, Box 30167, Nairobi.

MOROCCO: CENTRE DE DIFFUSION DOCUMENTAIRE DU B.E.P.I. 8, rue Michaux-Bellaire, Rabat.

SOUTH AFRICA: VAN SCHAIK'S BOOKSTORE (PTY) LTD.
Church Street, Box 724, Pretoria.

SOUTHERN RHODESIA:
THE BOOK CENTRE, First Street, Salisbury.

UNITED ARAB REPUBLIC: LIBRAIRIE
"LA RENAISSANCE D'ÉGYPTE"
9 Sh. Adly Pasha, Cairo.

ASIA

BURMA: CURATOR, GOVT. BOOK DEPOT, Rangoon.

CAMBODIA: ENTREPRISE KHMÈRE DE LIBRAIRIE
Imprimerie & Papeterie Sarl, Phnom-Penh.

CEYLON: LAKE HOUSE BOOKSHOP
Assoc. Newspapers of Ceylon, P. O. Box 244, Colombo.

CHINA:
THE WORLD BOOK COMPANY, LTD.
99 Chung King Road, 1st Section, Taipeh, Taiwan.
THE COMMERCIAL PRESS, LTD.
211 Honan Road, Shanghai.

HONG KONG: THE SWINDON BOOK COMPANY
64 Nathan Road, Kowloon.

INDIA:
ORIENT LONGMANS
Calcutta, Bombay, Madras, New Delhi
and Hyderabad.
OXFORD BOOK & STATIONERY COMPANY
New Delhi and Calcutta.
P. VARADACHARY & COMPANY, Madras.

INDONESIA:
PEMBANGUNAN, LTD., Gunung Sahari 84, Djakarta.

JAPAN: MARUZEN COMPANY, LTD
6 Tori-Nichome, Nihonbashi, Tokyo.

KOREA, REPUBLIC OF:
EUL-YOO PUBLISHING CO., LTD
5, 2-KA, Chongno, Seoul.

PAKISTAN:
THE PAKISTAN CO-OPERATIVE BOOK SOCIETY
Dacca, East Pakistan.
PUBLISHERS UNITED, LTD., Lahore.
THOMAS & THOMAS, Karachi.

PHILIPPINES:
ALEMAR'S BOOK STORE, 769 Rizal Avenue, Manila.
POPULAR BOOKSTORE, 1573 Doroteo Jose, Manila.

SINGAPORE:
THE CITY BOOK STORE, LTD., Collyer Quay.

THAILAND:
PRAMUAN MIT, LTD.
55 Chakrawat Road, Wat Tuk, Bangkok.
NIBONDH & CO., LTD.
New Road, Sikak Phya Sri, Bangkok.
SUKSAPAN PANIT
Mansion 9, Rajadamnern Avenue, Bangkok.

VIET-NAM, REPUBLIC OF:
LIBRAIRIE-PAPETERIE XUAN THU
185, rue Tu-do, B. P. 283, Saigon.

EUROPE

AUSTRIA:
GEROLD & COMPANY, Graben 31, Wien, I.
B. WÜLLERSTORFF
Markus Sittikusstrasse 10, Salzburg.
GEORG FROMME & CO., Spengergasse 39, Wien, V.

BELGIUM: AGENCE ET MESSAGERIES
DE LA PRESSE, S. A.
14-22, rue du Persil, Bruxelles.

BULGARIA:
RAZNOÏZNOS, 1, Tzar Assen, Sofia.

CYPRUS: PAN PUBLISHING HOUSE
10 Alexander the Great Street, Strovolos.

CZECHOSLOVAKIA:
ARTIA LTD., 30 ve Smečkách, Praha, 2.
ČESKOSLOVENSKÝ SPISOVATEL
Národní Třída 9, praha, 1.

DENMARK: EJNAR MUNKSGAARD, LTD.
Nørregade 6, København, K.

FINLAND: AKATEEMINEN KIRJAKAUPPA
2 Keskuskatu, Helsinki.

FRANCE: ÉDITIONS A. PEDONE
13, rue Soufflot, Paris (V^e^).

GERMANY, FEDERAL REPUBLIC OF:
R. EISENSCHMIDT
Schwanthaler Str. 59, Frankfurt/Main.
ELWERT UND MEURER
Hauptstrasse 101, Berlin-Schöneberg.
ALEXANDER HORN
Spiegelgasse 9, Wiesbaden.
W. E. SAARBACH
Gertrudenstrasse 30, Köln (1).

GREECE: KAUFFMANN BOOKSHOP
28 Stadion Street, Athens.

HUNGARY: KULTURA, P. O. Box 149, Budapest 62.

ICELAND: BÓKAVERZLUN SIGFÚSAR
EYMUNDSSONAR H. F.
Austurstraeti 18, Reykjavik.

IRELAND: STATIONERY OFFICE, Dublin.

ITALY: LIBRERIA COMMISSIONARIA SANSONI
Via Gino Capponi 26, Firenze,
and Via Paolo Mercuri 19/B, Roma.

LUXEMBOURG:
LIBRAIRIE J. TRAUSCHSCHUMMER
Place du Théâtre, Luxembourg.

NETHERLANDS: N.V. MARTINUS NIJHOFF
Lange Voorhout 9, 's-Gravenhage.

NORWAY: JOHAN GRUNDT TANUM
Karl Johansgate, 41, Oslo.

POLAND: PAN, Pałac Kultury i Nauki, Warszawa.

PORTUGAL: LIVRARIA RODRIGUES Y CIA.
186 Rua Aurea, Lisboa.

ROMANIA: CARTIMEX, Str. Aristide Briand 14-18,
P. O. Box 134-135, Bucureşti.

SPAIN: LIBRERIA BOSCH
11 Ronda Universidad, Barcelona.
LIBRERIA MUNDI-PRENSA
Castello 37, Madrid.

SWEDEN:
C. E. FRITZE'S KUNGL. HOVBOKHANDEL A-B
Fredsgatan 2, Stockholm.

SWITZERLAND:
LIBRAIRIE PAYOT, S. A., Lausanne, Genève.
HANS RAUNHARDT, Kirchgasse 17, Zürich 1.

TURKEY: LIBRAIRIE HACHETTE
469 Istiklal Caddesi, Beyoglu, Istanbul.

UNION OF SOVIET SOCIALIST REPUBLICS:
MEZHDUNARODNAYA KNYIGA
Smolenskaya Ploshchad, Moskva.

UNITED KINGDOM:
H. M. STATIONERY OFFICE
P. O. Box 569, London, S.E.1
(and HMSO branches in Belfast, Birmingham,
Bristol, Cardiff, Edinburgh, Manchester).

YUGOSLAVIA:
CANKARJEVA ZALOŽBA, Ljubljana, Slovenia.
DRŽAVNO PREDUZEĆE
Jugoslovenska Knjiga, Terazije 27/11,
Beograd.
PROSVJETA
5, Trg Bratstva i Jedinstva, Zagreb.
PROSVETA PUBLISHING HOUSE
Import-Export Division, P. O. Box 559,
Terazije 16/1, Beograd.

LATIN AMERICA

ARGENTINA: EDITORIAL SUDAMERICANA, S. A.
Alsina 500, Buenos Aires.

BOLIVIA: LIBRERIA SELECCIONES, Casilla 972, La Paz.

BRAZIL: LIVRARIA AGIR
Rua Mexico 98-B, Caixa Postal 3291,
Rio de Janeiro.

CHILE:
EDITORIAL DEL PACIFICO, Ahumada 57, Santiago.
LIBRERIA IVENS, Casilla 205, Santiago.

COLOMBIA: LIBRERIA BUCHHOLZ
Av. Jiménez de Quesada 8-40, Bogotá.

COSTA RICA: IMPRENTA Y LIBRERIA TREJOS
Apartado 1313, San José.

CUBA: LA CASA BELGA, O'Reilly 455, La Habana.

DOMINICAN REPUBLIC: LIBRERIA DOMINICANA
Mercedes 49, Santo Domingo.

ECUADOR:
LIBRERIA CIENTIFICA, Casilla 362, Guayaquil.

EL SALVADOR: MANUEL NAVAS Y CIA.
1a. Avenida sur 37, San Salvador.

GUATEMALA:
SOCIEDAD ECONOMICA-FINANCIERA
6a. Av. 14-33, Guatemala City.

HAITI: LIBRAIRIE "À LA CARAVELLE"
Port-au-Prince.

HONDURAS:
LIBRERIA PANAMERICANA, Tegucigalpa.

MEXICO: EDITORIAL HERMES, S. A.
Ignacio Mariscal 41, México, D. F.

PANAMA: JOSE MENENDEZ
Agencia Internacional de Publicaciones,
Apartado 2052, Av. 8A, sur 21-58, Panamá.

PARAGUAY:
AGENCIA DE LIBRERIAS DE SALVADOR NIZZA
Calle Pte. Franco No. 39-43, Asunción.

PERU: LIBRERIA INTERNACIONAL
DEL PERU, S. A., Casilla 1417, Lima.

URUGUAY: REPRESENTACION DE EDITORIALES,
PROF. H. D'ELIA
Plaza Cagancha 1342, 1° piso, Montevideo.

VENEZUELA: LIBRERIA DEL ESTE
Av. Miranda, No. 52, Edf. Galipán, Caracas.

MIDDLE EAST

IRAQ: MACKENZIE'S BOOKSHOP, Baghdad.

ISRAEL: BLUMSTEIN'S BOOKSTORES
35 Allenby Rd. and 48 Nachlat Benjamin St.,
Tel Aviv.

JORDAN: JOSEPH I. BAHOUS & CO.
Dar-ul-Kutub, Box 66, Amman.

LEBANON:
KHAYAT'S COLLEGE BOOK COOPERATIVE
92-94, rue Bliss, Beirut.

NORTH AMERICA

CANADA: THE QUEEN'S PRINTER
Ottawa, Ontario.

UNITED STATES OF AMERICA: SALES SECTION,
UNITED NATIONS, New York.

OCEANIA

AUSTRALIA:
WEA BOOKROOM, University, Adelaide, S.A.
UNIVERSITY BOOKSHOP, St. Lucia, Brisbane, Qld.
THE EDUCATIONAL AND TECHNICAL BOOK AGENCY
Parap Shopping Centre, Darwin, N.T.
COLLINS BOOK DEPOT PTY. LTD.
Monash University, Wellington Road, Clayton, Vic.
MELBOURNE CO-OPERATIVE BOOKSHOP LIMITED
10 Bowen Street, Melbourne C.1, Vic.
COLLINS BOOK DEPOT PTY. LTD.
363 Swanston Street, Melbourne, Vic.
THE UNIVERSITY BOOKSHOP, Nedlands, W.A.
UNIVERSITY BOOKROOM
University of Melbourne, Parkville N.2, Vic.
UNIVERSITY CO-OPERATIVE BOOKSHOP LIMITED
Manning Road, University of Sydney, N.S.W.

NEW ZEALAND:
GOVERNMENT PRINTING OFFICE
Private Bag, Wellington
(and Government Bookshops in Auckland,
Christchurch and Dunedin)

[6381]

Orders and inquiries from countries where sales agencies have not yet been established may be sent to: Sales Section, United Nations, New York, U.S.A., or to Sales Section, United Nations, Palais des Nations, Geneva, Switzerland

---

Printed in Hong Kong
63-18218–November 1963–3,325

Price: US$1.25 or equivalent in other currencies

United Nations publication
Sales No.: 63.II.F.8.
E/CN.11/631